CONFIDENTIAL SECRET

PHASE II FINAL REPORT

PROJECT CHIVE

Volume I

SYSTEM DESCRIPTION - SUMMARY

CHIVE/R-3-65

1 March 1965

DIRECTORATE OF SCIENCE AND TECHNOLOGY
OFFICE OF COMPUTER SERVICES

SECRET

CONFIDENTIAL

WARNING

This material contains information affecting the National Defense of the United States within the meaning of the espionage laws, Title 18, USC, Secs. 793 and 794, the transmission or revelation of which in any manner to an unauthorized person is prohibited by law.

Project CHIVE:

Phase II Final Report

Volume I

SYSTEM DESCRIPTION - SUMMARY

CHIVE/R-3-65

1 March 1965

010365

U.S. Central Intelligence Agency,
"Office of Computer Services.

DOC	/	REV DATE 07/1v/81	BY 018995
ORIG COMP —	OPI 63	TYPE //	
ORIG CLASS S	PAGES 121	REV CLASS C	
JUST 22	NEXT REV 2011	AUTH: HR 70-2	

TABLE OF CONTENTS

FIGURES

TABLES

Chapter 1.1.

SCOPE AND PURPOSE OF THE REPORT

This report represents the final product of Phase II
of Project CHIVE (System Design)--a joint OCS/OCR project
to study the information processing needs of the Agency
and to design and implement an EDP-based system reflecting
these needs.

The extensive documentation effort that was under-
taken to produce this report was meant to serve several
purposes:

- to inform Agency management of progress on the
  project and to seek its endorsement to proceed
  into Phase III - Initial System Implementation.

- to provide a sound mechanism for the detailed
  coordination between the designers of the system
  and the future operators of the system (the
  Office of Central Reference) which is vital if
  the functions and techniques proposed here are
  to be successfully implemented.

- to communicate the substance of the design to
  the rather large group of people coming from a
  variety of backgrounds which must participate in
  system implementation and to insure that they are
  all speaking the same language.

Effective communication of the concept and details
of a design effort in the complex area of Agency "information
processing" is a task comparable to the design task itself.
As a member of the OCR CHIVE Support Staff remarked in an

SCOPE AND PURPOSE
1.1.

early paper he wrote reacting to the CHIVE design concept, "Reading this paper is somewhat like jumping into water--you are surrounded no matter what your point of entry." There is no simple way to take one piece of the problem at a time, digest it, then go to the next one. Consequently, any organization of the material suffers from frustrating redundancy. The mode of presentation chosen here is roughly as follows.

What we are proposing is given in:

- Volume V, "System Organization, Functions, and Procedures," which is concerned with the elements of the system external to the mechanical tools which it will employ.

- Volume VI, "Document Delivery System," where the recommendations for document image storage and retrieval mechanisms are presented along with the analysis of alternative methods.

- Volume VII, "EDP System," where the proposed computer functions and equipment requirements are described.

How we propose to get started is described in Volume III, "Implementation Plan for the Initial System."

Why the design was undertaken and how we went about it is summarized in Volume IV, "System Requirements, Design Approach."

Implications for Agency management--benefits, costs, risks, alternatives, problems--are given in Volume II, "Management Summary."

SCOPE AND PURPOSE
1.1.

-2-

SECRET

In order to keep the size of the report within reasonable bounds, a conscious effort was made to restrict the text to a description of the proposed system, keeping philosophical points and arguments for or against alternative approaches to each problem area to a minimum. A discussion of rationale is included only where description would be inadequate without it or where we could do little more than state a problem. This accounts for the brevity of Volume IV.

No one volume of the report stands alone; liberal use is made of cross references--again, to keep the size down. However, the volumes are organized in sections of increasing detail so the reader can go as deeply into them as he desires without fear of missing a key point buried in the back. The casual reader should read at least through Volumes I and II as well as the early sections of Volumes III and IV. Evaluation of the proposed system requires careful attention to Volumes III and V. The system specialist, unfortunately, must face the task of reading the whole report.

One of the most significant tasks in Phase II was an extensive indexing experiment. Reference is made to it throughout the report; however, a detailed evaluation

SCOPE AND PURPOSE
1.1.

of the experimental results is not complete at this writing. They will be published in a supplement to Volume V.

Volume I is largely redundant with respect to all other volumes (except the Management Summary). The intent here is to distill the major <u>descriptive</u> elements of the rest of the report but to retain enough detail to make reading it informative and useful.

A bibliography of CHIVE Phase II documentation is included at the end of Volume I. A computer-generated index to the report will be published under separate cover.

It should be emphasized again that Volume I does <u>not</u> overlap with the Management Summary (Volume II). The former is a summary of the design product; the latter attempts to interpret this product with respect to risks, costs, benefits and the like. For the reader to get an overall picture of the project as it stands at the end of Phase II, he should devote equal portions of his time to each of the first two volumes.

<div align="right">
SCOPE AND PURPOSE
1.1.
</div>

-4-

Chapter 1.2.

SYSTEMS FUNCTIONS AND CAPABILITIES

The purpose of the proposed system is to provide
a wide range of information processing and reference
services to the Agency.

The most straightforward way to begin a discussion
of system capabilities is to say that the CHIVE* system
is to perform the functions now included within OCR.
Understanding what OCR does (and to some extent, how
it does it) is a prerequisite to understanding what is
proposed in this report. Even though the discussion is
sometimes put in the context of upgrading the functions
of OCR, little attention is given explicitly to a
description of the functions of that office as a whole.

CHIVE is not concerned with modifying the mission
of OCR (or of other central repositories) but with
how this mission might better be accomplished. Several

---

*The term "CHIVE" unfortunately has been corrupted in the
course of the project so that it now has three meanings:
(a) the name of the project, (b) the personnel team which
has been involved in the design of the system, and (c)
the name of the proposed system itself.

<div align="right">

FUNCTIONS AND CAPABILITIES
1.2.

</div>

-5-

past studies have analyzed specific aspects of OCR,
but Project CHIVE is the first concerted attempt to
study its functions within an overall framework.

Distilled to its simplest terms, OCR exists as a
centralized activity because:

- It can perform functions that an analyst in his normal routine cannot or should not do.

- It is an effective extension of his memory.

- It is an identifiable entity in the area of "information processing," the vital link between collection and production.

- Good management practice demands in some cases that large files of lasting interest can be best maintained as a centralized activity.

The proposed CHIVE system is an attempt to integrate
several (but not all) of the activities of OCR into one
system, applying advanced information processing techniques
and management tools to achieve a significant increase in
processing effectiveness.

1.2.1. SYSTEM FUNCTIONS

From one point of view, at least, CHIVE (or OCR)
is not a "system" at all. It is really a homogeneous
aggregate of functions carried on by an identifiable organi-
zation.

FUNCTIONS AND CAPABILITIES
System Functions
1.2.1.

-6-

The thing which gives identity to this organization is the bank of information which exists in many forms at its core. Associated with this bank of information are the services which can be performed, and the mechanisms needed to massage the data, add to it, and find new resources to exploit.

For convenience, the processing necessary to provide these services is divided here into two classes: document file processing and information file processing. The distinction between the two is not as as clear-cut as one might hope, but is sufficiently precise for descriptive purposes.

1.2.1.1. Document File Processing

This includes:

- A repository for positive intelligence material.

- Reproduction and distribution of documents.

- Document indexing and retrieval.

The principal recommendation here is to integrate the various document repositories, and to provide several points of view in accessing the material held by the system. This involves changes in the exploitation of a wide variety of sources, a reorganization of personnel

FUNCTIONS AND CAPABILITIES
System Functions
1.2.1.1.

in order to provide an environment for substantive
specialization, the use of standard sets of indexing and
retrieval aids, and improved data processing mechanisms
such as a computer and document image processing devices.

1.2.1.2. Information File Processing

This includes:

- Extraction and synthesis of elements from perhaps
  several document index records; that is, the
  process necessary to change from a document-
  oriented file to one which is organized around
  intelligence entities, such as personalities and
  facilities.

- The maintenance of reference aids for indexing
  and retrieval, such as indexing codes and
  dictionaries.

- The retrieval of information and production of
  reports from files.

- The inclusion of special-purpose files not
  produced in the course of normal processing,
  but which should be integrated in the search
  product.

Our recommendations here reflect a change in the
mechanism for OCR file processing, initially, and,
ultimately, a more significant change in the scope of
information files. That is, quite early in system
development, we hope to provide a flexible mechanical
structure applicable to most information file situations.
With this structure as a beginning, when applied to the

FUNCTIONS AND CAPABILITIES
System Functions
1.2.1.2.

-8-

data elements extracted during the normal input flow,

considerable power should be available to strengthen

OCR's ability to distill out significant intelligence

information from its diverse data base. CHIVE's

interest, then, is in growth potential--as yet

untested--away from a document retrieval orientation

and toward an information retrieval orientation.

1.2.1.3. Auxiliary Functions

A machine capability will provide assistance in

several other areas beyond the major file processing

functions:

- Assist in the production of OCR formal and
  informal publications, such as the Intelligence
  Publications Index and permuted title indexes.

- Production of reports for OCR management.

- Maintenance of file profiles and analyst profiles
  representing file holdings external to the system
  and areas of special analyst competence.

1.2.2. EXCLUSIONS

As indicated above, all of these functions--and

more--are currently being performed within OCR. The

significant exclusions from the CHIVE design are enumerated

below. These functions will affect (and be affected by)

the CHIVE system activities; indeed, some of them may be

FUNCTIONS AND CAPABILITIES
Exclusions
1.2.2.

-9-

worthy of detailed study now. They were excluded for
the moment because they are being adequately performed
now or would require an excessive amount of work to bring
them within manageable proportions.

- Liaison and administrative staff functions.

- Publications procurement and exploitation services.

- Dissemination services.

- Book cataloging and circulation.

- Miscellaneous special collections and ad hoc services.

1.2.3. SYSTEM CHARACTERISTICS

Figure 1-1 shows the basic elements of the system and
how they interact to produce the ultimate products.

The principal resource required, of course, is people
woven into an efficient organization and performing the
normal document and information analysis and clerical
functions. The anticipated workload on personnel and the
skills which they are expected to master are considerable.
The information analyst is expected to comprehend the
substance of the material and the questions he must
process, as well as understand the mechanics of indexing
and file manipulation. The approximate workload can be
summarized as follows (assuming an eventual total system):

- Indexing of 1.1 million documents per year, at
least 60% of these requiring some degree of content
analysis.

FUNCTIONS AND CAPABILITIES
System Characteristics
1.2.3.

-10-

Figure 1-1

SYSTEM ORGANIZATION
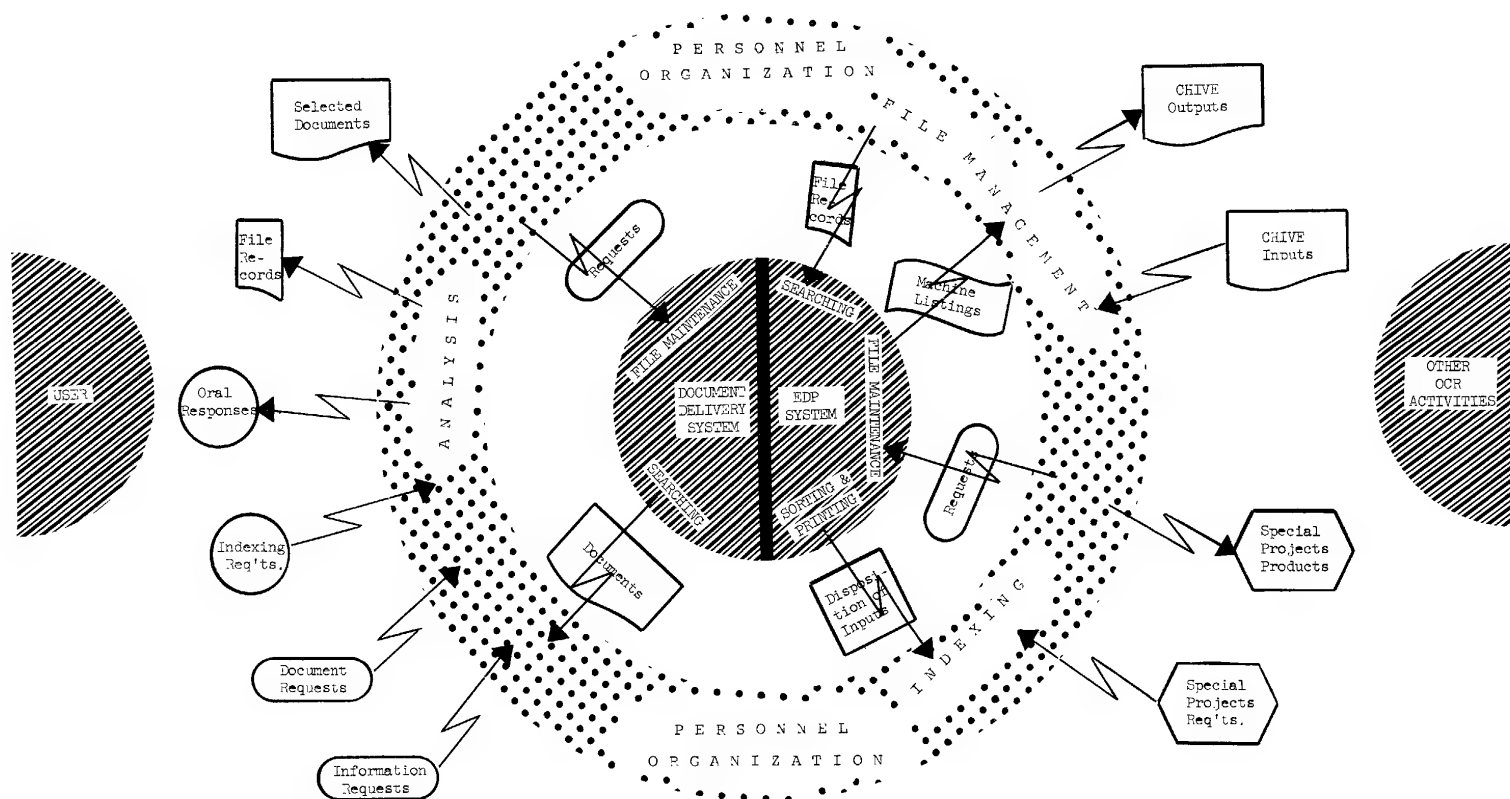
- Servicing 400 requests for document references,
  1,000 requests for specific information daily.

A so-called document delivery function has been
identified which includes the personnel and equipment
resources needed to photograph, store, retrieve, and
reproduce the images of the documents included in the
system; to maintain adequate management control of the
file; and to provide the necessary interface with the
computer system. This function will provide services
directly to the system customers as well as to elements
of the information processing organization.

Sometime in the useful life of most of the data in
the system, some representation of it will pass through
the computer. The computer with its associated equip-
ment and personnel will provide several facilities:

- It will act as a sorting, report formatting,
  and printing device.

- It will search the larger files in the system
  using search criteria expressed in a language
  which will provide a wide range of logical
  capabilities and which will be used directly
  by the information analysts.

- It will create and modify records in existing
  files.

- It will monitor all of its operations, checking
  for data validity, allocating its resources (both
  time and devices), and logging its activities.

The computer element of the system - an IBM
System/360, model 60, shared among several users -
will be capable of:

FUNCTIONS AND CAPABILITIES
System Characteristics
1 2 3

-12-

- Processing all index records and other input
  transactions into the files with a machine
  backlog of no more than 4 hours.

- Processing requests against computer files on
  a demand basis, giving a printed response within
  30 minutes (priority) or 4 hours (routine).

- Accepting inputs and delivering products via remote
  terminals of the computer (initially on an experi-
  mental basis only).

No unusual computer equipment is needed to perform

the CHIVE functions with the exception of bulk storage

devices (which would be added only as they are needed)

and a page reader that will transform typed input

material directly into machine form without the need

for an intermediate key punching step.

FUNCTIONS AND CAPABILITIES
System Characteristics
1.2.3.

-13-

Chapter 1.3.

SYSTEM ORGANIZATION

In general, the organizational philosophy of the
CHIVE system is to combine the required intellectual
talents of trained intelligence information analysts
with the processing and storage capabilities of the
computer. The source documents to be input to the
system, the necessary human functions to be performed
relative to these documents (i.e., reading, selecting,
indexing, querying and reporting), and the outputs to be
derived from the system are not significantly different
from those which characterize one or more elements of
the existing central reference operation. Only if the
proposed system is compared to an individual register
subsystem within the current OCR complex does the
contrast appear, and then only with respect to certain
features of the existing subsystem.

The responsibility for implementing a specific
organizational configuration must be left to those who
will direct the operation since there are a variety of
factors to be considered which are beyond the purview
of the system designer. To assist those, however, who

will be charged with this activity, it might be useful

to summarize the principal CHIVE organizational

recommendations in the context of the major functions

to be performed within the system.

1.3.1.   INPUT CONTROL AND CUSTOMER SERVICE

The CHIVE system would be built, largely, around

information analysts, organized (at the first level)

into some four or five geographic components.*

The information analyst would be responsible for

determining not only what documents entered the system

files but what data within these documents was captured

for retrieval purposes.

---

*It is our view that it is difficult to identify any
better way of organizing the input and retrieval activity
than by grouping the primary individuals involved by
geographic area. This approach loses the advantage of
source specialization in processing and poses the problem
of geographic overlap in document analysis and query
coordination. At the same time, it contributes to
standardization of vocabularies and procedures so important
in an all-source environment, and is in focus with customer
inquiries which normally relate to a particular geographic
region of the world. Thus, on balance, while it does not
overcome all operational problems that can be envisaged,
of all the alternatives considered it seems to come
nearest to meeting the system objectives.

<div align="right">

SYSTEM ORGANIZATION
Input Control
1.3.1.

</div>

-16-

The information analyst would also be responsible
for the selection and processing of data input to
information files required by customers, and would handle
all queries levied on the system.

Whether the information analyst should also specialize
by topic within area or by some class of intelligence
data (e.g., biographic, installation, etc.) remains a
moot point. CHIVE favors the former in the belief that
it would lessen the number of times a document would have
to be handled, but additional testing of both concepts
is desirable.

1.3.2. INDEX PREPARATION

The function of physically preparing the index records
to documents, including both the header (bibliographic)
as well as the content data descriptions, would be assigned
to special personnel operating in close communication with
the analytical components. The content indexers, like the
information analysts, would be subdivided by geographic
area and each would normally process the output of his
counterpart analyst or analysts.

Content indexers would each have a set of the
dictionaries and other vocabulary control tools pertinent
to his area of responsibility. In addition, a master set

SYSTEM ORGANIZATION
Index Preparation
1.3.2.

-17-

of other area dictionaries would be located within each
Content Indexing group for reference purposes. Content
Indexers would translate the items of data tagged by
information analysts into the codes and other descriptors
dictated by the vocabulary of the system.

Header data indexers would perform a function similar
to content indexing, but on the bibliographic elements
of a document. One group of header data indexers would
operate in a centralized mode, serving all geographic
components by header indexing, immediately upon receipt,
those documents for which CHIVE has a repository
responsibility. Other header indexers would be assigned
to each geographic organization to capture the necessary
bibliographic data pertaining to non-repository-type
documents which had been reviewed by information analysts
and selected for retention by the system. Header data
indexers can type their inputs in a form suitable for
processing by a page reader. However, a central pool of
typists will also be needed, operating at the system
level, to convert the majority of transcript sheets
received from Content indexers, as well as search requests
from information analysts, into the graphic quality
required.

<div style="text-align:right">

SYSTEM ORGANIZATION
Index Preparation
1.3.2.
</div>

-18-

1.3.3. IMAGE PROCESSING AND DOCUMENT FILE MAINTENANCE

Image processing is that activity conducted by the so-called "Document Delivery System," i.e., the microfilming and associated operations required to convert incoming documents to microimage form, as well as the reproduction of items retrieved from the document store for delivery to customers. Its principal interface is with the document store itself to which materials are passed after microfilming and from which it receives, in turn, items to be reproduced.

During the evolutionary development of the CHIVE system both the new and old system operators will require access to many of the same document collections. If the logistical problems are not too severe, it would seem advisable to co-locate all master document files in one general physical area to lessen the communication problem as well as make file maintenance operations more efficient. Similarly, because of the close relationship between the document files themselves and the image processing function, it is recommended that the latter be connected both physically and organizationally to the former.

SYSTEM ORGANIZATION
Image Processing & File Maintenance
1.3.3.

-19-

1.3.4.   MACHINE FUNCTIONS

The principal machine-related activities include:

- EAM personnel and equipment needed to input data
  to files not yet absorbed into the new system and
  to retrieve data therefrom. Assuming no conversion
  to an EDP storage medium, the latter, in particular,
  will necessitate the retention of an EAM facility
  for as long as the inherited files have value.

- Personnel needed to operate the new system, including
  associated input-output devices (e.g., the page
  reader), and the computer.

- System analysts/programmers (referred to in this
  report as EDP file analysts) who will develop
  and refine the machine operations to be performed,
  define new files to the system, etc.

SYSTEM ORGANIZATION
Machine Functions
1.3.4.

-20-

Chapter 1.4.

DATA BASE AND INDEXING

1.4.1. SOURCES TO BE EXPLOITED

Practically all sources of intelligence interest
will be reviewed for inclusion in the system, but the
OCR analysts will be selective regarding that information
which is actually indexed into the system. Based on
written selection criteria and knowledge of customer
interests, redundant reporting and reporting on subjects
or named objects of low level interest will be excluded.

The following sources will be reviewed for
exploitation:

- Photo Interpretation Reports (T/KH)

- COMINT

- Cables

- Open Sources, including original domestic and
  foreign publications, translations, and summary
  or title translations, ████████████████████████

25X1A

25X1A
25X1A

- Maps. Map Library exploitation will be incorporated
  into the system.

- Ground Photographs and Films. Graphic Register
  exploitation will be incorporated into the system.

- Finished Intelligence

DATA BASE AND INDEXING
Sources to be Exploited
1.4.1.

STATSPEC

- Raw Intelligence Reports--reports produced by DIA, State airgrams, OOB and CS reports, etc.

Table 1-1 shows the current volume of input from these sources. It also indicates our recommendation on the sources for which the system should assume repository responsibility and a gross indication of the indexing control to be exercised for each.

1.4.2. BASIC SELECTION CRITERIA

Selection criteria will depend on several factors: (a) the documents used and information needed by the analytic offices; (b) the all-source concept and organizational configuration thereof. These two factors have to be balanced against the manpower and resultant capability available for the operation.

There seems to be a consensus of opinion that several levels of indexing should be applied to the various categories of documents:

- Entire series to be indexed in depth

- Entire series to be rejected for depth indexing, but to receive header or bibliographic control.

- Entire series to be rejected completely.

- Specific documents within a series to be indexed in depth.

<div align="right">
DATA BASE AND INDEXING
Basic Selection Criteria
1.4.2.
</div>

25X1C

Selection of an indexing level for a particular
document category is contingent upon customer reaction and
acceptance, which determination requires discussion of
interest in series not covered now, and re-examination of
series presently covered. Customer interest will be
determined from previous surveys, from an examination of
present request patterns in OCR, and from a study of the
levels of control required for various documents or named
objects as reflected in the practice or experience of the
OCR analysts. Based on this knowledge, the CHIVE
information analyst will direct the indexer as to coverage
and depth, i.e., which personalities, which organizations,
and/or which subjects should be indexed. The CHIVE
Indexing Experiment has shown the need for title coverage
of most documents regardless of the level of indexing,
unless the document or series is completely rejected.
This includes title preparation for those types to be
selectively indexed which have no titles, e.g., non-CIA
cables.

DATA BASE AND INDEXING
Basic Selection Criteria
1.4.2.

1.4.3.   INFORMATION CONTROL CONCEPT

1.4.3.1.   Document/Information Retrieval

The system will provide combined information retrieval and document retrieval capability. Documents themselves will be at the heart of the system, with their index records providing access to them through content control. The index records will also be the base from which information files can be built. That is, in the process of indexing documents, facts about named things of intelligence interest will be extracted and stored. The approach will be to extract information about specific named objects, keep this information in the context of the document for document retrieval, and manipulate this information out of context for information retrieval. Summary records will be formed and maintained on select high-interest personalities, installations, and other finite subjects, but the creation of these records will be an analytic activity requiring the synthesis of index records and documentary information.

In addition to the index records, the indexer working aids will themselves be a source of answers to questions. For example, the Organization Identifier List will contain names of organizations, their locations, type of activity, etc.

DATA BASE AND INDEXING
Information Control Concept
1.4.3.1.

1.4.3.2. Manual Indexing

Automatic indexing is still largely experimental and is not sufficiently precise to meet most of the Agency's retrieval requirements. Automatic indexing techniques usually involve word frequency counts, assigning weights to high-frequency words, and storing these words as index terms. Other techniques include syntactic analysis, sometimes in conjunction with the above statistical process, It is obvious that these techniques could not be applied to an intelligence storage and retrieval system requiring high precision and recall, since much intelligence information is inferential and interpretive and requires analysis for high-quality indexing.

Human indexing therefore, with its recognized faults, is still superior to automatic techniques and is the only feasible system for CHIVE. However, some documents will require only title indexing, and in those cases automatic title-indexing techniques can be applied.

1.4.3.3. Subjects vs. Named Objects

Intelligence analysts have found that the "named objects"--e.g., installations, personalities, organizations-- most often provide the clues to resolving research problems.

> DATA BASE AND INDEXING
> Information Control Concept
> 1.4.3.3.

-26-

OCR request experience is an accurate reflection of this interest. We recommended, therefore, that these subjects receive the greatest emphasis; and, in view of OCR experience relating to the kinds of things users are interested in concerning named objects, we recommend that an increased number of attributes of named objects be brought under control. The latter are the elements of information which identify a named object, e.g., a person's address, organizational affiliation, etc. In-depth indexing of named-object attributes does not necessarily have to mean an equivalent increase in the volume of data indexed or in indexing time. Common attributes such as addresses, types of organizations, and products of an installation will be stored in indexer identifier lists; it will not be necessary to re-index this data when it is reported repetitively in documents.

We recommend that subject indexing, that is, the kind of indexing performed by the Intellofax System and the Subject/Commodity Section of the Special Register be continued at least to the present level, but on a broader data base to include important document series (e.g., ▮▮▮▮ **25X1C** reports) which are excepted today.

> DATA BASE AND INDEXING
> Information Control Concept
> 1.4.3.3.

1.4.4. INDEXING LANGUAGE

The basic element of the index language is a term
which consists of a tag and value. A tag is a three
character symbol which specifies the kind of entry
which follows the tag. For example,the tag PNO specifies
a persons name and the tag ONO specifies an organization
name. The function of the tag is to distinguish among
homographs and to expedite file organization. The value
is the index entry specified by the tag. Index terms are
written as follows:

(PNO) SMITH, JOHN

(ONO) CENTRAL INTELLIGENCE AGENCY

Fifty-nine tags are currently included in the CHIVE
indexing system and each tag represents an element of
information. The elements of information included in the
system are people, organizations/facilities, conferences/
meetings, places, and attributes of these so called named
objects as well as subject indexing of concepts, activities,
and commodities. Table 1-2 summarizes the basic elements of
information to be captured in the system. Table 1-3 shows
the elements for organization/facilities in more detail.

Some tags specify very precise elements, e.g., PDT =
Personality Travel Date; others specify very broad elements,

<div style="text-align: right">

DATA BASE AND INDEXING
Indexing Language
1.4.4.

</div>

SECRET

Table 1-2

ELEMENTS OF INFORMATION

Personalities	Name and Variants	Occupation
	Citizenship	Travel
	Education	Awards
	Affiliation	Location

Organizations/	Name and Variants	Subordination
Facilities	Function	Location

Conferences	Name	Type
	Sponsor	Location

Locations	Country	Post Box
	Province	Telephone
	Name and Variants	Cable Address
	Coordinates	Street Address

Bibliographic	Classification	Source
	Dissem. Controls	Language
	Evaluation	Countries
	Dates	Page Count
	Title	Cross references

SECRET

SECRET

Table 1-3

ELEMENTS OF INFORMATION;
ORGANIZATIONS/FACILITIES

Name   Translated or foreign language name,
                        previous names, abbreviations,
                        telegraphic code, identifier number

Nationality   Country code when different from
                        location

Subordination   Identifying number of parent
                        organization

Function   Coded functional type

Location   Country/province code, place name,
                        coordinates, cable address, post box
                        number, street address

Other   Coded "indicators" of other
                        information available on organization
                        facility; e.g., physical description,
                        facility security, status, descriptive
                        summary

e.g., RSC = Subject Index Code. In some cases, the
value is transcribed as it appears in the document; in
other cases it is reformatted, e.g., dates; and in other
cases it is taken from various supporting tools to
obtain consistency and is entered in standardized form.
Supporting tools include the Intelligence Subject Code
for subject indexing, identifier lists for organizations
and personalities, and a gazetteer for place names.

A system of linkage among index terms is employed
to avoid mismatches of terms or "false drops" on
retrieval. A "phrase" is a group of index terms which
are syntactically linked together, and which come from
the same document. The linkage is recorded as a unique
number applied to each term in the phrase. For example:

Linkage Indicator	Term
26	(PNO) SMITH, JOHN
26	(ONO) CENTRAL INTELLIGENCE AGENCY
26	(LAC) VIRGINIA, LANGLEY

To the system, these three terms with the same linkage
indicator are translated into the phrase (sentence),
"John Smith is affiliated with CIA in Langley, Virginia."

In addition to the content indexing, the standard
bibliographic elements (header data) will also be indexed

DATA BASE AND INDEXING
Indexing Language
1.4.4.

-31-

for each document included in the system. The complete
index record (content and bibliographic data) will be
available for searching either for content or
bibliographic terms separately or conjunctively.

A sample of a complete content index record is shown
in Figure 1-2. In this example, index terms have been
coded and linked to represent the following information:

25X1C
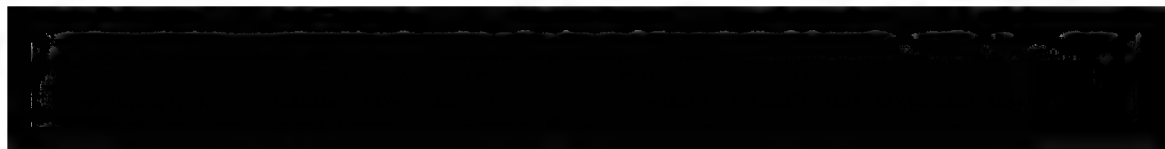
Individual phrases were formed and numbered 26-29. The
lines labeled A, B, and C contain entries which were
needed by more than one phrase. The letters appearing in
boxes on the right of phrases 26-29 indicate that the entry
on the referenced line is to be added to the phrase. Thus
phrase 26 also includes the entries on lines A and C.
In English, this phrase says:

Tag		25X1C Value
PNO	a person named	
POH	holds the position of	
ONO	in the organization	
LAC	Located at	
25X6	in the ▮▮▮▮ province	
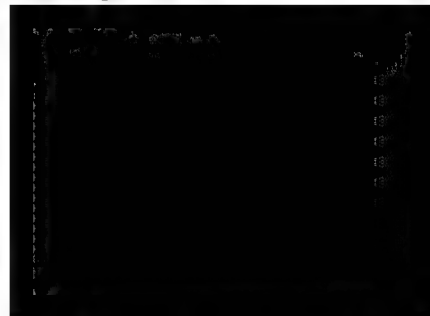
DATA BASE AND INDEXING
Indexing Language
1.4.4.

-32-

25X1C

Chapter 1.5.

SYSTEM FILES

The files described here are those which are identified
to the user--i.e., the CHIVE information analyst and,
perhaps ultimately, the research analyst. They are the
files he must be familiar with, if he is to take full
advantage of the resources of the system and exploit it
intelligently.

The total number of individual system files, including
old as well as new, might easily exceed a hundred. However,
it is possible to classify all the various files into
nine types, each with very distinctive functions and
properties. Table 1-4 shows the approximate size of
these files assuming growth to a full, integrated capability.

1.5.1. DOCUMENT INDEX FILES

These files contain all the raw document index records
in the system, including not only the complete index records
themselves but the directories to these records. The
documents referenced by these records may include any form
of information carrier--e.g., maps, photos, films or other,
and need not necessarily be readily accessible to the system.

SYSTEM FILES
Document Index Files
1.5.1.

-35-

Table 1-4

TABLE OF ESTIMATED FILE CLASS SIZES*

File Class ╲ System Development Period		1	2	3
CHIVE-built Document Index		75,000	200,000	525,000
Inherited Document Index		3×10^7	3×10^7	3×10^7
Summary		30,000	100,000	200,000
Unsynthesized		20,000	50,000	100,000
Vocabulary Control		150,000	350,000	750,000
Document Image		75,000	200,000	525,000

---

* These entries are made in terms of records. There may be duplicate machine records and listing records. Estimates are accurate to within one order of magnitude, based on current system design plans. CHIVE-built Document Index record estimates pertain to content-indexed documents only; documents solely header-indexed are not included. Inherited file estimates are germaine to Intellofax and the SR Detail File only, assuming a one-for-one, card-to-EDP-record type of conversion for the entire file. Summary file estimates are based on organization and personality summary files only.

1.5.2. VOCABULARY CONTROL FILES

These files are required to insure consistent entry of index terms (tag and value) into the Document Index Files and other system files. The principal function of these files is to reduce the synonym problem at search time. They include "identifier files" for named objects (which, like scope notes in a code schedule, help to distinguish one specific subject from another), code books, dictionaries, thesauri, and other authority lists.

1.5.3. UNSYNTHESIZED INFORMATION FILES

These files consist of select phrases or terms extracted from document index records or directly from the raw documents themselves. Such files would be built to facilitate retrieval where a substantial number of requests for the pertinent data can be anticipated on a continuing basis. Unlike Summary Information Files (see below), records in these files would often contain duplicative and/or contradictory information. Periodically, on demand,information in such files might be reviewed and added to the appropriate Summary Information Files.

1.5.4. SUMMARY INFORMATION FILES

These files are built either from records (or portions of records) in the Document Index Files, from records in

SYSTEM FILES
Summary Information Files
1.5.4.

Unsynthesized Information Files, or from the raw documents
themselves during or after input processing. The
distinguishing feature of these files is the fact that
they will ordinarily contain <u>evaluated</u>, <u>non-redundant</u>
data about named objects or events associated with named
objects. Named-object identifier files could be placed in
this file category, the only apparent difference being the
limited amount of historical data ordinarily found in such
files.

1.5.5. SPECIAL PROJECT FILES

The unique features of these files are as follows:
(a) the inputs to the files originate outside CHIVE;
(b) CHIVE actually acquires the files and not simply
"profiles" thereof; (c) additions or modifications to the
files can be anticipated; (d) the files do not use the
elements of information and/or vocabulary controlled in
CHIVE. Special Project Files may otherwise have the
properties of any of the file classes named above. These
files will be processed by CHIVE but maintained by CIA or
other Agency analysts.

1.5.6. REFERRAL SERVICE FILES

These files differ from Special Project Files in that
they are not substantive data files but rather descriptions

SYSTEM FILES
Referral Service Files
1.5.6.

-38-

SECRET

or profiles of files located outside the CHIVE system.
Referral Service Files will consist of profiles of analysts'
special fields of competence as well as files maintained
by analysts and/or information repositories external to
CHIVE. CHIVE will not maintain, or retrieve data from,
the substantive files themselves. It will simply inform
customers of those files (or persons) which might be
consulted relative to a given query.

1.5.7. DOCUMENT IMAGE FILES

These files contain documents stored by the CHIVE
system. From a functional point-of-view they include
"aspect" systems (where the index is stored separately
from the documents) as well as self-indexed document
files. Existing OCR document collections as well
as CHIVE-originated document repositories are encompassed
by this category. The storage media for such files will
include hard copy, various types of microimages, and
even digital storage in some instances. Similarly,
the categories of documents involved will differ widely
in size, shape, classification, and point of origin.
Physical storage of maps, ground photos, and films is not
considered a function of the CHIVE system. (Elaboration
of this point is given in Volume V.)

SYSTEM FILES
Document Image Files
1.5.7.

-39-

SECRET

1.5.8. MANAGEMENT DATA FILES

These files contain data collected on the activity
of the CHIVE system to (a) enable operational management
to evaluate the cost/performance figures of the system
and (b) to guide system designers in improving hardware
and software support. From the point-of-view of what
data is collected, most of the Management Data Files
will have to do with either system processing times
or processing volumes.

1.5.9. SYSTEM PROCESSING FILES

These files are used to support the system in
processing data. Most such files will be organized in
table form enabling values to be obtained from arguments.
Examples would include a file of legal tags and other
error correction files, a decode file which would convert
codes into clear text for display to a reader, intermediate
files which exist only temporarily during the processing
of a transaction, working storage files, etc. These files
are largely internal to the CHIVE EDP System, and the
information analyst need not interact with them in any
direct way.

For each of the file categories listed above, a
second-level categorization may be required, i.e., one

SYSTEM FILES
System Processing Files
1.5.9.

or profiles of files located outside the CHIVE system.
Referral Service Files will consist of profiles of analysts'
special fields of competence as well as files maintained
by analysts and/or information repositories external to
CHIVE. CHIVE will not maintain, or retrieve data from,
the substantive files themselves. It will simply inform
customers of those files (or persons) which might be
consulted relative to a given query.

1.5.7. DOCUMENT IMAGE FILES

These files contain documents stored by the CHIVE
system. From a functional point-of-view they include
"aspect" systems (where the index is stored separately
from the documents) as well as self-indexed document
files. Existing OCR document collections as well
as CHIVE-originated document repositories are encompassed
by this category. The storage media for such files will
include hard copy, various types of microimages, and
even digital storage in some instances. Similarly,
the categories of documents involved will differ widely
in size, shape, classification, and point of origin.
Physical storage of maps, ground photos, and films is not
considered a function of the CHIVE system. (Elaboration
of this point is given in Volume V.)

SYSTEM FILES
Document Image Files
1.5.7.

DUPL. PG.

1.5.8. MANAGEMENT DATA FILES

These files contain data collected on the activity of the CHIVE system to (a) enable operational management to evaluate the cost/performance figures of the system and (b) to guide system designers in improving hardware and software support. From the point-of-view of what data is collected, most of the Management Data Files will have to do with either system processing times or processing volumes.

1.5.9. SYSTEM PROCESSING FILES

These files are used to support the system in processing data. Most such files will be organized in table form enabling values to be obtained from arguments. Examples would include a file of legal tags and other error correction files, a decode file which would convert codes into clear text for display to a reader, intermediate files which exist only temporarily during the processing of a transaction, working storage files, etc. These files are largely internal to the CHIVE EDP System, and the information analyst need not interact with them in any direct way.

For each of the file categories listed above, a second-level categorization may be required, i.e., one

SYSTEM FILES
System Processing Files
1.5.9.

-40-

which classifies CHIVE files from the point-of-view of
their <u>origin</u>. These classes are three in number.

1.5.10. CHIVE-BUILT FILES

These are the files built by and for the CHIVE system
either from new inputs or through the conversion of existing
OCR files to the format and vocabulary of CHIVE. These
files will be continually updated as part of the regular
processing cycle.

1.5.11. INHERITED FILES

These are the files originally established by the
various OCR systems which it was not found possible to
integrate with new CHIVE files. Such files will include
records in hard copy as well as machine language. In
some instances these files may be transferred to another
storage medium (e.g., magnetic tape) if querying and
output can thereby be improved. Similarly some existing
machine-readable files may be restructured and interrogated
in the vocabulary of a single CHIVE language. Neither of
these changes, however, implies true conversion to the
CHIVE system. Another significant difference between
these files and CHIVE-Built Files is that while both will

SYSTEM FILES
Inherited Files
1.5.11.

-41-

be used by the CHIVE information analyst, no additions
will be made to the Inherited Files once the CHIVE System
is fully operational.

1.5.12. SUPPLEMENTAL FILES

In this class are the files not built or maintained
by CHIVE, nor inherited from OCR, but which contain data
functionally useful to CHIVE as a secondary source of
information. All Special Project Files (see above) fit
this category, as do reference aids of various kinds
(e.g., Who's Who compilations, gazetteers, commercially
published indexes, etc.) obtained from external sources
and left essentially in the form in which they were
received.

SYSTEM FILES
Supplemental Files
1.5.12.

Chapter 1.6.

SYSTEM FLOW

1.6.1. DOCUMENT INPUT FLOW

The information is received primarily in the form
of documents; however, index records to maps, photographs,
and films will also be included in the system, as will
certain machine-language data prepared on contract (but
under CHIVE control) by external organizations (e.g., the
Library of Congress). Graphics and maps will continue to
flow to GR and the Map Library Division (ML) through
their existing acquisition channels. The only significant
change in their operations will be that they will employ
the CHIVE vocabulary in their indexing or cataloguing
operations, and will transmit a copy of their index
transcript sheets to CHIVE for conversion into machine
readable form and entry into the Master Document Index
File. CHIVE will return to them a printed version of
their index records for entry into their manual files,
when desired.

Documents selected by the information analyst which
are available in machine language and have a formatted

SYSTEM FLOW
Document Input Flow
1.6.1.

-43-

header and title (e.g., SI Teletype) will bypass
indexing and transcription steps and go, in their
machine language versions, directly to the EDP System
where the necessary conversion to CHIVE format will
be performed. The hard copy versions of the documents
will be sent simultaneously to microfilming for
processing into the microimage store (Master Document
Image File). Other machine language receipts,
consisting of abstracts of foreign scientific and
technical literature, bibliographic records, and formatted
information extracts pertaining to named-object data
appearing in open sources, may likewise be input directly
to the computer.

Following preparation of the index record (a function
normally performed by humans except where only a limited
retrieval capability seems required), the index will be
converted to machine storage with the aid of a page
reader and placed in a random access device, ultimately
the IBM/System 360 Data Cell Drive. The information
storage capacity of one Data Cell Drive will
allow us to accommodate the content of an estimated
600,000 index records (the actual storage capacity is
400 million characters of information), and there is no

SYSTEM FLOW
Document Input Flow
1.6.1.

-44-

practical limit on the number of modules that could be provided. The same device would be used to hold the directory to the index records themselves, i.e., a list of the terms which appear in the index records and, for each term, the record and phrase number(s) containing the term. This will obviate the need to examine every index record in the file to see if it contains the terms sought.

Most textual documents will be converted to microfilm and then stored. Two storage systems are recommended in this report: 35mm aperture cards (containing up to 8 images per aperture) or packed microfiche (sheet microfilm records containing up to 60 letter-size pages on each microfiche). Documents in excess of a certain page limit as well as documents of poor image quality will be kept in hard copy. Maps, films, and photos will continue to be stored in the conventional manner in the physical repositories where they are now located.

Whether the 35mm aperture card or microfiche storage system is chosen, the document images would be filed in motorized card files but would be retrieved and refiled manually. Assuming 10 million documents were to be

SYSTEM FLOW
Document Input Flow
1.6.1.

stored on site, the estimated floor space required for a
packed microfiche system would be an area approximately
30' x 60'. Output from either the hard copy document
or microimage files would consist of paper copies. The
integrity of the document collection will be maintained
such that none of the master microimages, or original
documents if filed only in hard copy, will leave the file
except for photoduplication or hard copy printing.

1.6.2.   DOCUMENT RETRIEVAL

The retrieval process will begin with a customer
external to CHIVE originating a request for data either
on a form designed for this purpose, by telephone contact,
or by personal visit to the system. He will be put in
touch with an information analyst working on the geographic/
topical area of concern. The information analyst will be
familiar with the current reporting, having screened
incoming documents to determine what should be indexed,
and will also have had extensive training in both the
indexing vocabulary, the logical files available within
the system, and the query language required to conduct
the computer search.

SYSTEM FLOW
Document Retrieval
1.6.2.

-46-

SECRET

After ascertaining the clearance level of the customer, the degree of sensitivity desired in the search, and the heterogeneity of the document base to be explored (e.g., "search document and photo indexes, but not maps or films"), the information analyst (assuming a machine search is required) will translate the request into a set of commands using the formal language developed by CHIVE. To prepare the necessary search criteria he will consult the various Vocabulary Control Files in order to derive the proper terms on which the search would be conducted.

Having determined what descriptors to employ in the search, he will detail the logic and priority of the search, indicate the files to be searched, and define the output format required.

If the date span of the request encompassed the period prior to the initiation of the CHIVE system (which will be the normal case for several years), the information analyst may be required to take one or more of the following additional steps:

- Examine inherited hard copy files of cards or documents co-located with his organization component.

<div align="right">

SYSTEM FLOW
Document Retrieval
1.6.2.

</div>

SECRET

- Request the retrieval of hard copy records (e.g., MIRA, one-name cards, etc.) from the system's centrally-located, master document collection.

- Consult with other information analysts familiar with the contents, vocabularies, and record formats of machine files inherited by CHIVE and obtain their assistance (where required) in preparing the special request forms to interrogate said files.

The formulated machine requests will be typed and sight verified, and then transmitted to the Page Reader via the pneumatic tube system. For those records to be passed against the EDP files, the computer will check for such things as the completeness of the request statement and validation of the terms composing the query. All requests will then be queued for processing against the pertinent Inherited and CHIVE-Built Files.

Searches of unconverted EAM files will be conducted as at present, with the output taking the form of existing machine listings which cite documents, personality dossiers, installation numbers, or photo accession numbers relevant to the request. For files converted to EDP and the CHIVE-built Master Document Index File, the product of the search will be a list of the document control numbers which satisfied the search criteria, the complete "hit" index records or select elements thereof, or the count of

SYSTEM FLOW
Document Retrieval
1.6.2.

the number of documents which matched the search prescription, without the records themselves. All codes appearing in the records would be translated into clear text for ease of understanding by the information analyst and customer (if the latter also reviews the listing directly).

In some cases the output records may, themselves, answer the request. If so, the retrieval activity will end with the information analyst transmitting the desired information by mail or phone to the customer.

When the index record output is satisfactory but in itself does not supply the answer sought, the information analyst may order the pertinent documents from the Document Delivery System before transmitting the results of the search to his customer for review. Graphics and map index records uncovered during the initial search will be transmitted to the customer who may order these items himself.

The information analyst may be asked to respond to the inquiry by phone, memorandum, completion of a customer's response form, or by the preparation of a narrative report (e.g., a biographic summary or a piece of finished biographic intelligence). In this case, he would supply information rather than documents, which would necessitate a more sophisticated analysis and synthesis of the materials at hand.

SYSTEM FLOW
Document Retrieval
1.6.2.

Lastly, the information analyst may update certain of his identifier records, as well as dossier files, to reflect the results of his analysis or send a marked copy of his report (if it deserves retention) back through the input process for indexing and storage in the Master Document Image File.

1.6.3.   INFORMATION FILE BUILDING, MAINTENANCE, AND RETRIEVAL

As has been pointed out, the CHIVE system, like the existing central reference operation, will require a variety of dictionaries and other support tools (given the general title of Vocabulary Control Files in this report). In addition, it will maintain substantive files of information either in unsynthesized or summary form. These files, unlike the Master Document Index records, will require continual maintenance, i.e., the deletion of obsolete data as well as the correction or addition of data in existing records.

1.6.3.1. Vocabulary Control File Maintenance

Vocabulary Control Files will be consulted by indexers in order to select approved terms or codes. If the indexer finds no suitable entry or if it is erroneous or incomplete, he will specify the change to be made to

SYSTEM FLOW
File Building, Maintenance, Retrieval
1.6.3.1.

-50-

the file in question using a portion of the same command
language employed in the retrieval of records from the
Master Index File

A dictionary editor will be responsible for reviewing
all changes made to this specific vocabulary Control File.
He will insure that the proposed transaction is legitimate
and proper.

The transcript sheet will be converted to machine
language by the page reader and fed to the EDP System for
updating the pertinent machine files. A record of the
changes made will then be printed out in the various
arrangements required, and returned to the indexers. The
frequency of preparation of these printed supplements to
master listings, as well as the frequency with which the
master listings themselves will be rerun, will vary
depending on the number of changes occurring over a given
period of time.

1.6.3.2. Information File Processing

As indicated previously, formatted information files
consisting of logical data units either in unsynthesized
or summary form may be initiated either: (a) by analysts
external to the CHIVE system having a pressing and

SYSTEM FLOW
File Building, Maintenance, Retrieval
1.6.3.2.

-51-

continuing need for the retrieval of select facts (as
distinct from documents) or (b) by CHIVE information
analysts reacting to the accumulation effect of specific
request patterns. Requirements of this nature, since
they will increase both the human and machine processing
burden, should be reviewed by managers at the branch or
higher level to determine the anticipated load on the
system and its capacity to respond to same.

Accepted requests for the establishment of such
files will be assigned to one or more information analysts
who will consult with an EDP file analyst. The latter
will be thoroughly familiar with the internal operations
of the EDP System and, in particular, the method used to
establish new digital files. He will design the format
and record structure of the machine file required by the
information analyst and see to it that the file is actually
established.

In general, the approach of the information analyst
will be to use the document retrieval capability to help
build the required information files. This will be
particularly true in the building of an Unsynthesized
Information File. In this case, the data requested is

SYSTEM FLOW
File Building, Maintenance, Retrieval
1.6.3.2.

probably already reflected in the content of document

index records (i.e., the UIF would be built directly from

rearranged elements of index records). Where this is

indeed the case, the information analyst will periodically

direct the computer to take such action by calling for

the appropriate standing query and record generation job

to be run.

Summary Information Files, on the other hand, will

require more activity on the part of the information

analyst since they will consist of evaluated, summary

records about named objects or events. These can be

generated only by the analysis of the output from an

Unsynthesized Information File, of the Master Index File,

or by the processing of the incoming documents themselves.

Assuming this file is to be built from data in an

Unsynthesized Information File, the information analyst

will review the listed product from the latter, comparing

it with a listing of any records already stored in the

Summary Information File. If he decides to add new data,

delete what was there, or replace old information with new,

he will prepare a File Maintenance Transcript Sheet.

This form will follow the usual path to typing, thence to

SYSTEM FLOW
File Building, Maintenance, Retrieval
1.6.3.2.

-53-

SECRET

the Page Reader, and finally to the EDP System for computer processing.

The retrieval of data from either Unsynthesized or Summary Information Files might be initiated for a variety of reasons:

- To provide a listing of changes in the master file in order to update the information analyst's printed version of the file.

- To provide a listing of the complete master file either for reference use by the information analyst or for periodic publication and distribution to interested customers.

- To search, in response to a customer's request, for a specific fact or correlation of facts which could not be readily derived by human browsing of the printed records.

The retrieval process will be virtually the same as that followed in the retrieval of document index records (using the same retrieval language). Schedules can be set up for the levying of standing queries which would cause the listing of all or a portion of a file on a periodic basis without any action being required on the part of the responsible information analyst.

SYSTEM FLOW
File Building, Maintenance, Retrieval
1.6.3.2.

Chapter 1.7.

COMPUTER INTERFACE

The EDP portion of CHIVE will perform the following
functions:

- Build and maintain files

- Create sub-files from existing files

- Search files and retrieve data from them

- Display data

The techniques chosen to implement these functions
provide a built-in flexibility that will also allow
revisions in the definition of the content and structure
of CHIVE-built files.

An integral part of the proposed EDP system is a
command language that allows these types of manipulation.
It is recognized that "unlimited" flexibility is possible
if the user could be persuaded to use machine language
directly. More practically, a set of commands is provided
that permits personnel other than programmers to use the
EDP system.

The language designed for the CHIVE system allows the
user to direct the performance of the four functions

COMPUTER INTERFACE
1.7.

mentioned above. Full use of the commands requires good knowledge of the indexing procedures, logic, and knowledge of the content and structure of the records and files to be manipulated. It is planned that information analysts and some content indexers will be trained to use the language.

The responsibilities concerned with defining new files and modifying existing file definitions will be assigned to the EDP file analyst. The EDP file analyst must be trained to a level similar to that of a programmer since he must be able to specify files to the system, initiate jobs for the machine operations personnel and participate in subsequent check-out.

1.7.1. CHIVE DATA ELEMENTS AND THEIR LOGICAL STRUCTURE

The CHIVE language is designed for use with a generalized logical data structure which is user oriented. All CHIVE EDP records are similar in their logical structure to the structure of the document index record as it is developed by the analyst in the document indexing process. Moreover, the user can think in terms of a hierarchic structure of terms, phrases, and records in each file. The structure is identical for a file of

COMPUTER INTERFACE
Data Elements and Logical Structure
1.7.1.

-56-

document index records and for information files of
any functional type (e.g., vocabulary lists, summary
files, and unsynthesized files). Figure 1-3 shows the
structural similarity of a document index record and a
summary file record. This similarity will facilitate
the transfer of data from a document orientation to
an information orientation. As far as the user is
concerned, the records in all of these files will be
serially searched and scanned.

1.7.2. STRUCTURE AND FUNCTIONS OF THE COMMAND LANGUAGE

A transcript sheet has been designed for the direct
entry of the processes which the information analyst wishes
to have the EDP system perform. He describes his job as a
sequence of commands such as COPY, DELETE, PRINT, etc.
For each of these commands the information analyst specifies
a number of parameters--what to copy, delete, print, etc.

1.7.2.1. Test Conditions

In order to provide a wide range of search criteria,
the analyst fills in "test conditions" which specify the
kinds of tests he wishes to make in the course of the
execution of the commands. There are two types of test
conditions--simple and complex. A simple test condition

COMPUTER INTERFACE
Structure & Functions, Command Language
1.7.2.1.

-57-

Figure 1-3

RECORD STRUCTURES

Document Index Record

| record control phrase | document # and other bibliographic data |

| phrase 1 | person 1 date of birth organization name affiliation date |

| phrase 2 | person 1 travel location purpose of travel travel dates |

| phrase 3 | commodity shipped "from" location "to" location |

| phrase 4 | organization name function location parent organization |

| phrase 5 | persons 2,3, and 4 leader appearance date location |

etc.

Personality Summary Record

| record control phrase | personality name and personal data |

| phrase 1 | educational institute 1 dates attended degree |

| phrase 2 | educational institute 2 dates attended degree |

| phrase 3 | organization affiliation 1 dates of affiliation |

| phrase 4 | organization affiliation 2 dates of affiliation |

| phrase 5 | travel location 1 travel dates |

| phrase 6 | travel location 2 travel dates |

etc.

is the basic logical expression written on one line of the
transcript sheet. Complex test conditions reference
other tests (either simple or complex), so that the true
or false result of a complex test condition is a function
of the true or false result of its constituent tests.

The most common simple test condition is one in
which a tag and a value is specified and a match is sought
within the records of the file being processed. In
addition simple test conditions can be specified which
ask for something other than a perfect match. For example,
does the record or phrase currently being examined have
the given tag with a value that is less than or greater
than the value specified in the test condition. A span
test condition would be true if the record or phrase
currently being examined has the given tag and has a value
which is between two specified values. A scan test
condition would specify a combination of required characters
and "don't care" characters.

The true or false result of a complex test condition
is a logical function of the true or false results of
specified simple tests. For example, a complex test
condition might specify that at least two of its four
constituent tests must be satisfied. Another feature

COMPUTER INTERFACE
Structure & Functions, Command Language
1.7.2.1.

of complex test conditions is the ability to specify whether
it must be satisfied within a single phrase of the record
or whether it could be satisfied anywhere in the record
without regard to phrase linkage.

1.7.2.2. Commands

The various commands available in the command
language are briefly described below.

The COPY command copies those records, phrases or
terms which meet the search criteria onto an output file.
These data elements are copied from a named input file.
The PRINT command prints a named input file (generally the
COPY output file) according to a named output format. The
EXPLODE command is similar to the COPY command except that
each phrase or term in the named input file which meets
the specified search criteria is copied as an individual
record on the named output file.

When it is necessary to edit or modify "hit" data
elements, an EXTRACT . . . WRITE command sequence must
be used. APPEND, DELETE, REPLACE and MERGEPhrases commands
can be included in the sequence to modify the retrieved
data. The EXTRACT command extracts (into a temporary work
area) records, phrases and terms which meet the search
criteria. This extracting occurs one record at a time.

COMPUTER INTERFACE
Structure & Functions, Command Language
1.7.2.2.

-60-

A succeeding WRITE command will write out the modified
contents of the work area into a named output file. An
intervening APPEND command appends data to the phrase(s)
or term(s) in the work area which meet the search
criteria. The DELETE command deletes data elements which
meet the specified search criteria. The REPLACE command
replaces the data elements in the work area record which
meet the search criteria with data specified in the
command. The MERGEPhrases command operates on the single
extracted record in the temporary work area; all phrases
within the record with a common value for a specified
merge tag are merged into a single new phrase.

There are three computational commands--TALLY,
ACCUM and COMPUTE--which perform computational operations
on CHIVE data elements. TALLY counts the number of
records, phrases or terms in a named input file which
meet the specified search criteria. ACCUM accumulates
the sum of the numeric valued terms in a named input file
which meet the specified search criteria. COMPUTE
performs arithmetic operations (add, subtract, multiply
and divide) on CHIVE data elements and on user-defined
values.

COMPUTER INTERFACE
Structure & Functions, Command Language
1.7.2.2.

-61-

The MERGERecords command causes all records in a
named input file, which contain a common term(s) to be
physically co-located in a named output file.

1.7.2.3. Jobs

A variety of combinations of these commands can be
specified as "jobs" for the computer to process. The
do-it-yourself flexibility of the language gives the
information analyst considerable power beyond the capability
he needs to retrieve records from machine stored files:

- He can build an Unsynthesized File directly
  from document index records, retaining only the
  data he needs and structuring it to suit his
  purposes.

- Files built in one job can be saved by the EDP
  system and manipulated by the Information Analyst
  as separate entities at any time thereafter--using
  the same command language.

- Complex retrieval jobs requiring access to several
  files can be run.

- In a single job, the product of one search can be
  used directly as search criteria for another
  without manual intervention.

Some examples of retrieval and file building processes
which the information analyst can specify directly in the
command language are given below. Each example would be
handled as a single job.

> COMPUTER INTERFACE
> Structure & Functions, Command Language
> 1.7.2.3.

-62-

25X6

- List all index records for documents containing information about ███████████████████████ in order by date of information.

25X6

- List the names and locations of all radar facilities within the coordinate square defined by ████████, 25X6 ████████.

25X6

- List the index records containing information about the ██████████████████████████ List the records showing the names of people associated with the plant first, followed by those containing no personality references. Count the number of records in each category.

25X6

- Extract individual phrases from index records which indicate the organizational affiliations of ██████████████████████████ Form a file which contains one record for each such person showing all of his organizational affiliations.

25X6

- List the people who have been associated with both ██████████████████████████████████ done even though tying a person to both organizations may require correlation of information from two or more independent records).

- List all known participants and dates of all conferences held in 1963 which were attended by personnel associated with the ████████████ 25X6 (Note: here the personnel linked to the radar plant might be stored in several records and the conference data in several other records, or perhaps even in a different file.)

1.7.2.4. File Maintenance

The second major consideration of the user is to build and maintain files. The usual file maintenance operations are provided. They are:

> COMPUTER INTERFACE
> Structure & Functions, Command Language
> 1.7.2.4.

- Adding new data to a file

- Changing existing data

- Deleting existing data

The information analyst can control the file maintenance
operations in either of two ways. The first way is the
usual one of specifying a unique record identification and
then having the desired maintenance performed on that record.
The second way is to specify logical conditions that could
qualify a single record or many records within a file
for the specified maintenance operation. For example,
it may be desired to change the names of all factories
named the Stalin Works to Big Brother Industries. In
such a case it is only necessary to set up the test
condition with a REPLACE command. The desired changes
are made without requiring that the user know in advance
the unique identifications of all of the records involved
in the transaction.

1.7.3. FILE DEFINITIONS AND THE EDP FILE ANALYST

The CHIVE command language allows manipulation of
data in existing files and also permits a way of creating
sub-files which can in turn be processed by the EDP system.
These features directly concern the information analyst.

COMPUTER INTERFACE
Definitions & the EDP Analyst
1.7.3.

-64-

SECRET

The tasks and procedures associated with changing file definitions and adding new files to the system are the responsibility of the EDP file analyst. The CHIVE EDP programs are controlled by external description of the data files to be processed. The data descriptions taken collectively are called File Format Tables. Each table describes a file and its constituent elements. If it is desired to process files other than those currently defined it is necessary to add new table descriptions to those already in existence.

The File Format Tables contain all the information about an item that is required to process it. Included are the terms allowed in a record, term groupings, which terms are used as identifiers, addressing parameters, occurrence data, how stored, and content legality parameters. Extensive revisions can be made to the tables. In addition to adding new files, terms can be added to or deleted from an existing file. Legalities can also be changed. It is important to note that <u>revisions of this type do not require any maintenance to the EDP programs</u>.

COMPUTER INTERFACE
Definitions & the EDP Analyst
1.7.3.

-65-

Chapter 1.8.

DOCUMENT DELIVERY SYSTEM

1.8.1.   DEFINITION OF THE DOCUMENT DELIVERY SYSTEM

The Document Delivery System may be generally
defined as that segment of the total CHIVE system which
deals with the input, storage, and recovery of identified
documents. It is self-contained in the sense that it will
not be electronically interconnected with the computer-
based indexing system and will not have a computer
capability of its own. The system will have repository
responsibility for textual documents only, with maps and
graphics retained elsewhere. The input will be, primarily,
hard copy documents from a variety of sources and will
range from poor to high quality printing. The documents
are to be locatable by some identification number such
that the file can be interrogated directly by the user
(Counter Service Requests) or indirectly through a search
of the computer index (Query Requests). In response to
either type of request, the system must furnish a usable
replica copy of the document master, since the master is
not to be circulated outside the file.

<div align="right">
DOCUMENT DELIVERY SYSTEM
Definition
1.8.1.
</div>

The system corpus has the potential of growing
to an extremely large size over a period of time. The
magnitude of the system is defined within this volume
at two points on the project growth curve referred
to as the:

- Initial System

    - Input--100,000 documents/year

    - Request rate--500 requests/day

- Total System

    - Input--1,000,000 documents/year

    - Request rate--5,000 requests/day

A maximum repository volume of 10,000,000 documents has
been assumed as a long range design goal. Although the
specific hardware to be used does not have to be
identical throughout, some upward capability must be
demonstrated to allow for transitional growth from the
Initial to the Total System.

1.8.2. RECOMMENDATIONS

The two systems which were found to be most favorable
from the standpoint of economics and performance are the
Packed Microfiche and the Filmsort Aperture Card. The
systems are described briefly below.

DOCUMENT DELIVERY SYSTEM
Recommendations
1.8.2.

-68-

Packed Microfiche: Microfiche are sheet microfilm records considered here as conforming to a 105mm x 148mm (4 in. x. 6 in.) standard which can contain up to 60 letter-size pages. Document images are recorded on 105mm roll microfilm by means of a 'step-and-repeat' camera which automatically places consecutive exposures in a matrix format (6 x 12) with the upper row reserved for recording eye-visible identification information. Documents are allocated on the microfiche such that (a) multiple items may be recorded on each microfiche, (b) each new item would begin a new row with an eye-visible identification number in the left most column, and (c) no item containing less than 60 pages shall 'spill over' onto a second microfiche. The original silver shall be duplicated onto diazo roll film for backup file purposes. Cut diazo microfiche will be filed in motorized files in the sequence recorded. On demand, selected microfiche are enlarged to hard copy by means of a Xerox Automatic Microfiche Printer.

Filmsort 2000dx: This system utilizes 35mm aperture cards as its basic storage medium. With the introduction of the 3M Filmsort 2000dx camera, a fully processed aperture card containing up to eight page images can be

DOCUMENT DELIVERY SYSTEM
Recommendations
1.8.2.

-69-

produced. Backup records are produced from the original
by means of a Copy-Reproducer. File copies are stored
in motorized card files in Document Control Number
sequence. Selected items are pulled from the file and
hard copy is produced on the 3M Quadrant Printer.

Filmsort 1000d: Developed by the Minnesota Mining
and Manufacturing (3M) Co., the Filmsort 1000d is a
combination camera-processing unit which provides a
finished 35mm aperture card within 60 seconds after the
original document is placed under the camera. This
single piece of equipment allows aperture cards to be created
without the separate film processing and mounting operations
performed in most existing aperture card systems. Duplicate
backup files are created by an automatic aperture card
copier. The cards are filed in motorized card files and,
on demand, may either be duplicated at the file using
the UNIprinter 086 approach (similar to that for the
16mm aperture card system) or be removed from the file
to a central automatic hard copy enlarger.

DOCUMENT DELIVERY SYSTEM
Recommendations
1.8.2.

Chapter 1.9.

EDP SYSTEM

The <u>initial</u> CHIVE EDP system has the following basic

goals and aims:

- Demand processing of transactions with user assigned priorities, through use of the multiprogramming features of the IBM Operating System/360. This provides minimum turn-around times for high priority transactions

- Inputs are primarily from a centralized optical page reader

- Experimental remote communication capabilities will be provided

- Many information requests will be answered from periodic printouts from machine stored files

The <u>total</u> CHIVE EDP system will have the following

additional goals and aims:

- Remote input terminals will enable information analysts (and, ultimately, users) of the CHIVE system to query and maintain data files directly on an up-to-date basis. This will in many cases supplant the answering of information requests from periodic printouts.

- A greater number of data files will be put "on line," i.e., continuously readable, in order to reduce the turn around time of maintenance and retrieval transactions.

EDP SYSTEM
1.9.

-71-

1.9.1.   DATA PROCESSING FUNCTIONS AND HARDWARE

The term "EDP system" represents both logical and physical entities. Logically, the CHIVE EDP system consists of a combination of files and processing functions as outlined in Figure 1-4. Physically, the CHIVE EDP system consists of various components of the OCS computing center as outlined in Figure 1-5.

The CHIVE EDP system comprises the following major information handling and retrieval functions:

- File structure definition

- File creation

- File maintenance

- File querying and information retrieval

- Processing of retrieval data

- Report format definition

- Output report generation

Subject to continuing review during Phase III and Phase IV, and specifically evaluated prior to growth

25X1A   beyond the ███████████████, the following hardware components will be shared with other agency jobs processed by the OCS computer:

- Optical page reader

- Central processing unit - IBM/360 Model 60

                    EDP SYSTEM
                    DP Functions & Hardware
                    1.9.1.

-72-

Figure 1-4

OVERALL VIEW OF CHIVE EDP SYSTEM

Figure 1-5

EDP HARDWARE SYSTEM

Card Reader
. Input

Page Reader
. Input

Printer
. Report generation

Central Processer
. executive control
. data manipulation for retrieval & maintenance

Core Storage
. Programs and data being processed

Console
. Input/Output

Fast Random Access Storage

Bulk Random Access Storage

Magnetic Tapes

. Program Library
. System tables
. Temporary storage for transactions in process

. Data Files

. Data Files
. Temporary storage for transactions in process
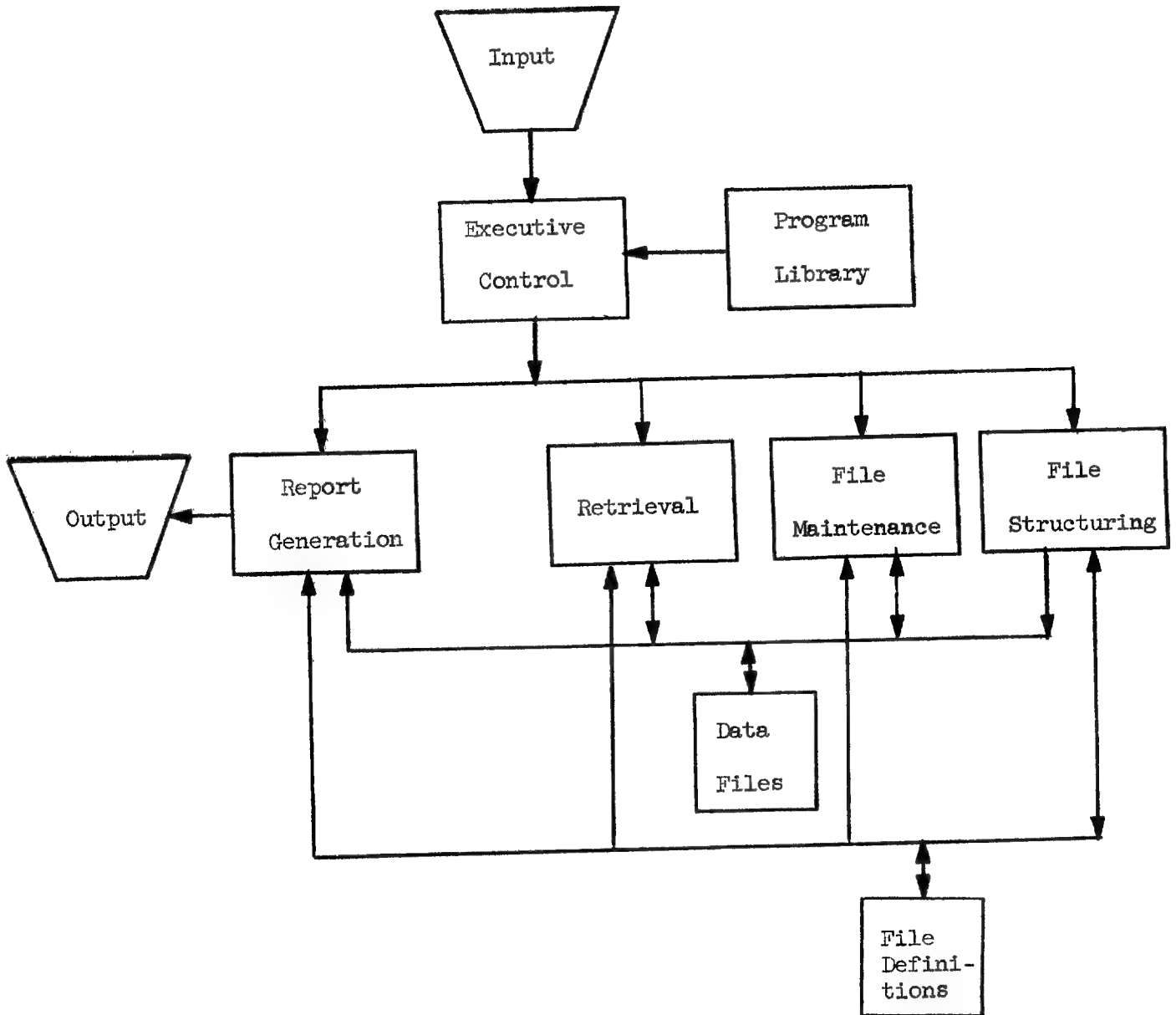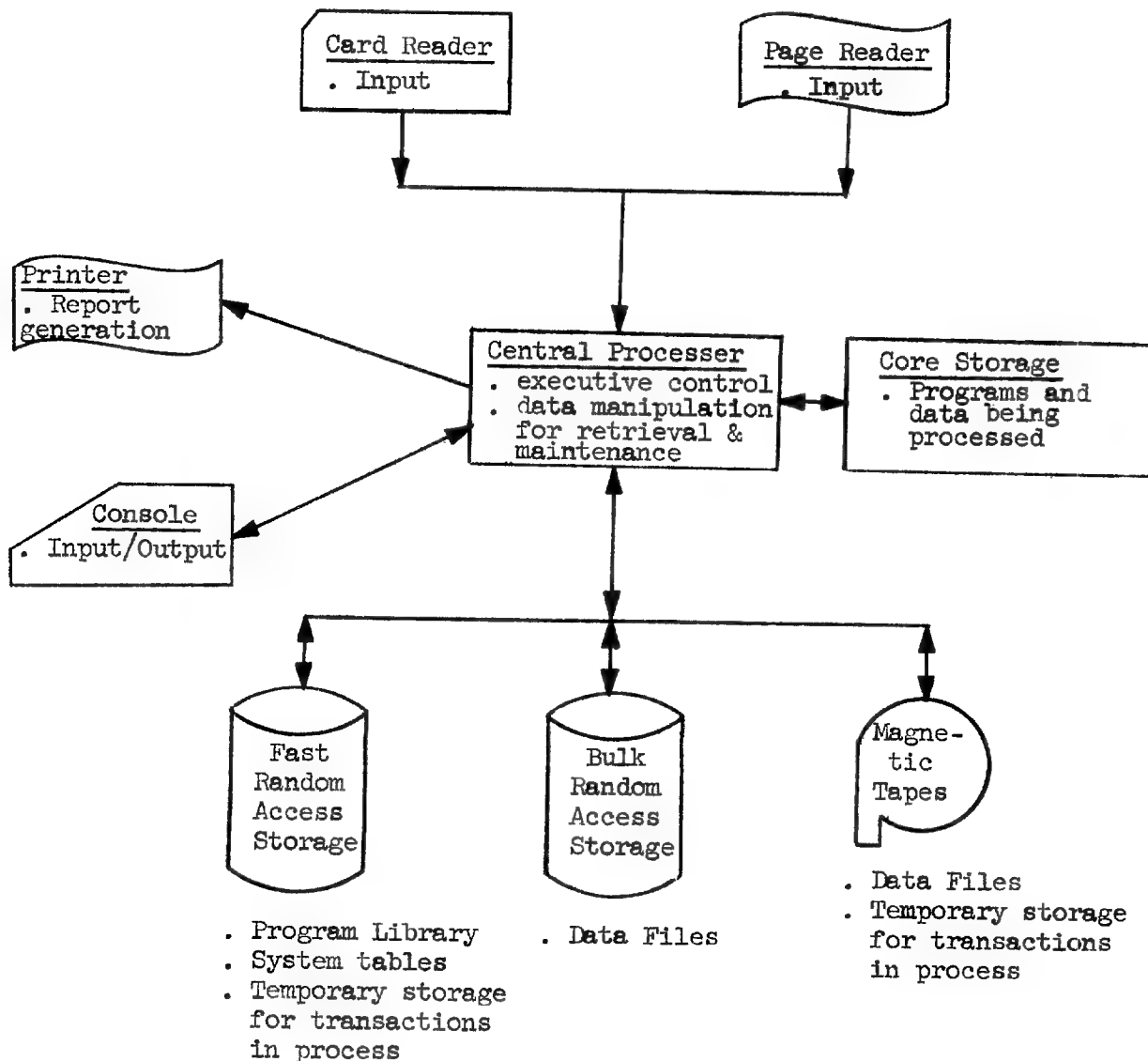
-74-

- Core storage (a) High speed - 512,000 bytes
  (b) Bulk core - 1 million bytes

- Input/output data channels

- IBM 2321 data cell drive

- IBM 2302 disk storage device

In addition the following auxiliary storage devices are devoted exclusively to CHIVE use:

- Magnetic tapes

- IBM 2311 diskpaks

- IBM 2321 data cells

1.9.2.   EDP FILES

Chapter 1.5. describes CHIVE files from a user or non-EDP point of view. A different classification scheme is more appropriate as far as the EDP system is concerned.

1.9.2.1.   System Data Files

Practically all CHIVE user files, i.e., those described in Chapter 1.5., are classified as system data files, because of the common CHIVE record structure. CHIVE system data files include: document index record files, identifier files, dictionary files, unsynthesized information files, and summary information files.

EDP SYSTEM
EDP Files
1.9.2.1.

-75-

1.9.2.2. Directory Files

When a system data file is stored on a direct access storage device, such as a disk file or data cell, a companion directory file may be used to facilitate accessing and searching of the data file. The directory file will contain a record for the majority of terms in the records of the data file. Each term record will list the identification of the phrases in each record of the system data file which contain the particular term. If a search is to be made for phrases containing two specific terms, for example, the two directory records for these terms would be compared to determine which document/ phrase numbers are shared by both records. The index records themselves would then be retrieved.

1.9.2.3. System Processing Files

These files are required for effective maintenance and retrieval from the system data files. They contain standing job definitions, report format definitions, file format tables, and input format definitions.

1.9.2.4. Executive Control Files

These files will be maintained by the IBM Operating System/360 (see below). These files contain the CHIVE

EDP SYSTEM
EDP Files
1.9.2.4.

-76-

program library, activity logs, storage maps, and tables
to control buffering of computer input and output.

1.9.3.   SYSTEM EXECUTIVE CONTROL

An integrated set of programs called Operating
System/360 will be provided with the IBM System 360
which will control the execution of all machine processes
in such a way as to make maximum use of the facilities
available at any particular time and to provide an effective
multiprogramming capability. Multiprogramming involves the
concurrent processing of many jobs within a computer, as
opposed to the more traditional approach of sequentially
processing jobs one at a time. Multiprogramming executive
control in the Operating System/360 primarily involves
allocation of the scarce computing resources among competing
concurrent jobs. Major resource allocation activities
include:

   - System input and output

   - Central processing unit

   - Core storage

   - Data management

EDP SYSTEM
System Executive Control
1.9.3.

-77-

1.9.4. FILE ACCESSING AND RECORD SEARCHING

Files must be accessed and individual records searched for two main reasons:

- To maintain the records in a file by (a) creating new records, (b) appending new phrases or terms to records (c) deleting phrases or terms from a records, and (d) replacing phrases or terms with new data

- To retrieve information which has been stored in the records of a file

CHIVE will utilize two of the standard Operating System/360 file accessing methods:

- Sequential access method

- Indexed sequential access method

For the sequential access method the records in the file are related to each other by position, as on a tape. The indexed sequential access method applies to any direct access device, such as a disk file, and permits both sequential and random accessing. The cylinders and tracks on which the records are stored are maintained by OS/360 in index tables associated with the file.

Certain large and very active system data files will be accessed through the use of directory files, as noted previously. The term posting records in the directory files will themselves be accessed with the OS/360 indexed sequential method. The posting list

EDP SYSTEM
File Accessing & Record Searching
1.9.4.

-78-

in the directory files will be maintained automatically
at the same time that their associated system data files
are maintained.

EDP SYSTEM
File Accessing & Record Searching
1.9.4.

-79-

SECRET

Chapter 1.10.

IMPLEMENTATION OF THE INITIAL SYSTEM

The initial CHIVE system is intended to be a microcosm of the proposed total system. It will model and test, as far as is possible, major system and organizational goals. It is designed to provide a point from which a total system can develop in an evolutionary manner. While a number of limitations (such as incomplete data base and coordination problems with extant OCR components and files) will constrain the initial system, it should provide sufficient experimental information to enable Agency management to decide whether to proceed with implementation of a total system.

It is recommended that implementation of the system be done on an incremental basis, beginning with █████████ 25X6

25X6   ██████ For initial implementation purpose, this country was selected because:

- It has a managable document volume.

IMPLEMENTATION OF INITIAL SYSTEM
1.10.

- It is a country of sufficient intelligence significance to generate more than average consumer interest.

- Its documentation gives a sampling of major topic areas--i.e., political, scientific and technical, military, economic.

STATSPEC

- Its documentation includes the full range of information carriers to be encompassed by CHIVE processing; e.g., ████ open literature, cables, finished intelligence, raw intelligence, Comint, T/KH, etc.

- There are available personnel familiar with the geographic area and its documentation.

25X6

The input flow of ███████████ documents or items which can be anticipated for the initial system breaks down roughly as shown in Table 1-5.

Table 1-5

Estimated Annual Volume of ███████████

25X6

Open literature (███████ items)	14,879
STATSPEC ████████	16,900
Raw Intelligence Reports (documents)	12,759
Finished Intelligence (documents)	1,045
SI/T/KH (documents)	20,750
Miscellaneous	14,160
Total	80,493

25X1A

1.10. IMPLEMENTATION OF INITIAL SYSTEM

DUPL. PG.

Chapter 1.10.

IMPLEMENTATION OF THE INITIAL SYSTEM

The initial CHIVE system is intended to be a microcosm
of the proposed total system. It will model and test, as
far as is possible, major system and organizational goals.
It is designed to provide a point from which a total system
can develop in an evolutionary manner. While a number of
limitations (such as incomplete data base and coordination
problems with extant OCR components and files) will
constrain the initial system, it should provide sufficient
experimental information to enable Agency management to
decide whether to proceed with implementation of a total
system.

It is recommended that implementation of the system
be done on an incremental basis, beginning with ▮▮▮▮▮▮▮ 25X6

25X6 ▮▮▮▮ For initial implementation purpose, this country
was selected because:

- It has a managable document volume.

IMPLEMENTATION OF INITIAL SYSTEM
1.10.

-81-

DUPL. PG.

- It is a country of sufficient intelligence significance to generate more than average consumer interest.

- Its documentation gives a sampling of major topic areas--i.e., political, scientific and technical, military, economic.

STATSPEC
- Its documentation includes the full range of information carriers to be encompassed by CHIVE processing; e.g., ████ open literature, cables, finished intelligence, raw intelligence, Comint, T/KH, etc.

- There are available personnel familiar with the geographic area and its documentation.

25X6
The input flow of ███████████ documents or items which can be anticipated for the initial system breaks down roughly as shown in Table 1-5.

Table 1-5

Estimated Annual Volume of ███████████   25X6

Open literature ██████████ items)	14,879
STATSPEC ████████	16,900
Raw Intelligence Reports (documents)	12,759
Finished Intelligence (documents)	1,045
SI/T/KH (documents)	20,750
Miscellaneous	14,160
Total	80,493

25X1A

IMPLEMENTATION OF INITIAL SYSTEM
1.10.

Based on a study of current activity in OCR, a one-year
projection of the estimated distribution of China requests
processed by existing components (which would eventually
be processed in the initial CHIVE system) is shown in
Table 1-6.

Table 1-6

Estimated Request Volume on ▓▓▓▓▓▓▓▓▓▓▓▓   25X6

Intellofax	178
Foreign Installations Branch	352
Special Register	320
Biographic Register	1,128
Total	1,978

(Graphics Register was unable to describe its request volume
by geographic area.)

1.10.1.   INITIAL ORGANIZATION

Initial system implementation has been projected over
an eighteen-month period. To provide the organizational
framework in which to achieve the initial system goals,
it is proposed that an experimental ▓▓▓▓▓▓▓▓ be   25X1A
established. The branch should be structured to perform
the functions intended for the initial CHIVE operational

IMPLEMENTATION OF INITIAL SYSTEM
Initial Organization
1.10.1.

-83-

component, and personnel should be added as required tasks are undertaken and needed skills identified. The test branch should be brought to operational strength around implementation month 10, and should move into an experimental indexing and training phase. (See timetable below.) Full system testing and operational simulation could begin in

25X1A

month 19, and the █████████████ could operate thereafter in parallel with current OCR operations until a decision could be made on the readiness of the initial component to assume operational responsibilities.

To assist in the implementation of the test component, and to provide logistical assistance and operator advice to OCS, the Phase II CHIVE Support Staff (CSS)* should be enlarged during month 1 to a five-man team consisting of four substantive analysts (suggested representatives: DD, FIB, SR and BR) and one OCR support programmer who should have a knowledge of MD or SR maintained EAM authority files. Two additional OCR support programmers should join the staff at

---

*After the report drafting was under way, OCR established a systems analysis staff whose main function, initially, is assisting in implementation of the test branch.

> IMPLEMENTATION OF INITIAL SYSTEM
> Initial Organization
> 1.10.1.

SECRET

a later time. Personnel from GR and FDD could be committed
on an ad hoc basis to work on inclusion of graphic indexes
and on the problem of obtaining machine-readable input

25X1A

from ███████████████

25X1A

Both the CSS and ██████████████████ would be
slotted against OCR's T/O. It is recommended, however,
that the CSS be based in OCS/Development Division where
work on the major system design problems and on development
programming is assigned.

1.10.2. IMPLEMENTATION TIMETABLE

A timetable showing the major milestones and suggested
implementation tasks is given below:

Month 1   Phase II work and Phase II plan review completed
           FDP equipment ordered (IBM System/360 Model 60)
           Secure specifications of Operating System/360
           Enlarged CSS on board
           China Test Branch nucleus formed
           Begin building indexer aids
           Begin specification of information and summary files
           Begin definition of input/output procedures
           Program design underway
           Review of indexing procedures begins

Month 2   Document Image System selected and ordered
           Analysis of OCR/CHIVE indexing experiment completed
           Specifications for page reader completed; page
             reader selected and ordered
           Study of Map Library and Graphics Register inputs
             begins

                          IMPLEMENTATION OF INITIAL SYSTEM
                          Implementation Timetable
                          1.10.2.

-85-

SECRET

25X1A

Month 3	Forms design completed Begin outlining dissemination procedures
Month 4	███████████████ T/O identified EDP equipment (IBM 360/Mod 30) on site Review of indexing procedures completed Program design completed Development programming begins
Month 5	Program unit testing begins
Month 6	Support file design completed
Month 7	Detailed design procedures for Document Image System completed
Month 8	Document site selected and prepared
Month 9	Document Image System installed Job sheets written for CTB Support files ready Indexing procedures established System test plan completed System test plan development started
Month 10	Dissemination procedures completed CTB brought to full operational strength Indexer training begins
Month 11	Initial information files fully defined Basic operating System/360 available (software) Page reader delivered and checked out Start experimental indexing
Month 12	360/Mod 60 equipment installed
Month 13	System operating procedures completed
Month 14	System test development completed
Month 15	Indexing tests completed Program unit testing completed

IMPLEMENTATION OF INITIAL SYSTEM
Implementation Timetable
1.10.2.

Graphics and Map Library inputs ready
Definition of interaction with OCR (projects,
  inherited files)

Month 16　Secure full Operating System/360
Final program checkout beings
Document dissemination started
Graphic and map inputs begin

25X1A

Month 18　Detail procedures for customer access to ███████
Branch
Customer training beings

Month 19　Operational testing begins

1.10.3.　FUNCTIONS OF THE INITIAL ORGANIZATION

The conclusion of the Phase III effort should have

modeled an initial CHIVE component which will be ready

for testing. The first testing stage, beginning in

month 19, will be a simulation of actual operational

experience--a shakedown, in a sense, of the final

configuration derived from the experimentation of the

preceding months. Some further experimentation, adjustment

of programs, and refinement of procedures, will carry over

into the initial testing stage. As far as possible, however,

the initial component will act like a fully operational OCR

component, i.e., it will receive, analyze, and index

documents, forward index records to the computer system,

IMPLEMENTATION OF INITIAL SYSTEM
Functions of Initial Organization
1.10.3.

SECRET

build files, and respond to queries, The major difference

25X1A between the ████████████ and a "live" OCR component will be
that it will not assume actual operational responsibility
to respond to queries, since these will be borrowed, or
captured at the time of receipt, from the various OCR
registers. It will also use the old OCR files maintained
by the registers.

1.10.4. PHASING IN OF NEW AREAS

Since the planned system would increment by geographic
25X6 area, ████████████████ could be the second element to be
added to the initial CHIVE component. The developing

25X6 ██

East Asia. At this point, assuming that system concepts
are working out as planned, the major priority area, the
25X6 ████████████████████. Eastern Europe might be the next
point of incrementation, followed by the Near East and
Africa, Western Europe, and Latin America.

IMPLEMENTATION OF INITIAL SYSTEM
Phasing in of New Areas
1.10.4.

Chapter 1.11.

COMPARISON WITH OTHER INTELLIGENCE SYSTEMS

The computer is beginning to play a significant role in the processing of non-numeric data in several intelligence environments. The design of the CHIVE system has drawn heavily from the experience gained in the development of intelligence information processing systems. Several of these are described briefly below. Each has one thing in common--the use of computers. They vary considerably, however, in the extent and philosophy of computer usage, in the functions that the systems are to perform, and in the data base processed.

1.11.1.   CIA SYSTEMS

25X1A

CIA Systems
1.11.1.1.

is used for handling micro-images of documents. Peripheral equipment associated with the latter device provides field offices and Headquarters components with aperture card sub-files.

The major keys for document retrieval are the names referenced in a document. These are annotated by the originator with review and additions made by an information specialist. A comprehensive name table is used in retrieval to cope with variant name spellings. This table, in machine language, will be employed as a retrieval aid in CHIVE.

Comparison with CHIVE: Both systems contain a large data base requiring the use of random access storage devices. Both have machine-stored information files employing generalized formats, and both keep index files referencing documents separate from the documents themselves. Because of these similarities, CHIVE should be able to use many of the vocabulary and information coding standards developed by ▇▇▇▇▇ However, there are significant differences in system purposes and the user population. CHIVE must be able to get at its stored data from almost any point-of-view (facilities,

25X1A

COMPARISON WITH OTHER SYSTEMS
CIA Systems
1.11.1.1.

25X1A

25X1A

commodities, trade, subjects--as well as personalities), but can afford to deliver a less precise product in many instances. ███████ on the other hand, must concentrate its effort on coverage of CI biographics. Further, the operational environment of DD/P demands close control of document circulation which ███████ (and RID) provides through a document locator system--rather than wide, parallel dissemination which is more suitable to the analytic environment.

1.11.1.2. Automatic Language Processing (ALP) System

The ALP system now under development, is a special-purpose computer configuration which will assume some processing functions in the Agency's exploitation of foreign language publications. The heart of this computer system is a large capacity disc which will store dictionaries and computer programs to be operated on by an associated "Lexical Processor." Two basic functions will be performed by this system initially:

- rudimentary translation of Russian, where the input will be paper tape punched from Cyrillic text. The machine output will be put in publishable form by post-editors.

- transcription of text to English from recordings produced by stenograph machine operators. Those operators will key the material directly from

COMPARISON WITH OTHER SYSTEMS
CIA Systems
1.11.1.2.

ardio tapes produced by translators from most
foreign language texts (except Russian). The
machine output will be mats ready for printing
after manual proofreading and correction.

Comparison with CHIVE: The ALP System is concerned
with the linguistic aspects of foreign language texts of
intelligence interest, while CHIVE is concerned with
providing retrieval "handles" for these texts (among
others). In the CHIVE concept, the foreign publications
exploitation function is similar to SIGINT data reduction--both
are "pre-processing" functions which must be performed on data
before it is handled by CHIVE.

The unusual computing techniques in table look-up
and symbol manipulation techniques used in the ALP
system will continue to be studied for their applicability
in CHIVE. When ALP is tied to the OCS IBM System/360,
some experiments in the use of ALP hardware as a table
look-up device for CHIVE will be performed.

1.11.1.3. SANCA (Security Automatic Name Check Activity)

25X1A

This system is similar to the basic ███████ activity
in that it is concerned with biographic data. The major
distinction is that SANCA will process data files (one
person per record) for the Security Records Division of OS.

COMPARISON WITH OTHER SYSTEMS
CIA Systems
1.11.1.3.

Each record will contain a reference to a security file

number. Access to the file will be by name (as in ███████, 25X1A

but will include machine access to attributes as well (such

as citizenship or occupation). "On-line" requests and

answers provided via remote terminals is a long-range goal.

Comparison with CHIVE: The file organization in

SANCA will be by name, but because of the need to access

the file by other terms, search problems similar to

CHIVE's must be solved. A major SANCA activity will be

updating file records. In this sense, the SANCA file will

be of the Summary Information Files class defined by CHIVE.

1.11.1.4. OCS Applications

The Intelligence Branch of the Applications Division

of OCS provides direct support to Agency analysts who may

need computer capability for any of several reasons:

building, editing, and printing files of intelligence data;

manipulating statistical data; format conversion; file

searching; etc. The number of files and the activities on

them have increased to the point where some degree of

programming generalization and file consolidation has

become feasible. As new programming requirements

arise, existing routines from an extensive library

are evaluated--and in most cases will satisfy at least

a part of the requirement. As an example of file

COMPARISON WITH OTHER SYSTEMS
CIA Systems
1.11.1.4.

consolidation, an Automated Target Information System has been implemented and presently includes three CIA files, two NSA files, and one SAC file concerned with collection targets.

Comparison with CHIVE: The CHIVE EDP objectives closely parallel the evolution of ad hoc applications in the Intelligence Branch--both deal with a variety of intelligence data (primarily non-numeric), both seek to provide a flexible EDP structure for file manipulation and retrieval, and both are dedicated to supporting the same kind of customer. The major differences are in the specificity of support to the analyst and in the place in the intelligence cycle where the data is captured. The Intelligence Branch supports individual analysts or components through tailor-made products derived from data of special interest to them. CHIVE, on the other hand, is concerned with capturing data of use to a significant portion of the total Agency analyst population and controlling it (at least initially) as reference keys to documents. In CHIVE parlance, the Intelligence Branch deals with Special Project files.

COMPARISON WITH OTHER SYSTEMS
CIA Systems
1.11.1.4.

As CHIVE moves toward a flexible information processing capability, the possible overlap of missions will have to be examined closely. A basic issue is when--both in time and in scope--does an automated analyst file move out of the special project and into the central reference domain (or vice versa)? Secondly, will the EDP capability required in both areas ultimately be almost identical? Regardless of the inefficiencies that may or may not ensue, the Agency must eventually provide a clarification to the analyst who otherwise would be bewildered by a seemingly ambiguous choice of services to tap.

1.11.2. DIA SYSTEMS

A number of computer activities in intelligence data processing are under way at DIA. Three major systems will be discussed here.

1.11.2.1. Unifile System

The Unifile system is designed as a computer data handling structure within which certain existing intelligence community files are to be placed. The

COMPARISON WITH OTHER SYSTEMS
DIA Systems
1.11.2.1.

system uses conventional computer equipment, but has
flexible record formats and command structures which
will accomodate a variety of inputs. The objective is
to convert some 3 million existing records pertaining to
55 community files to Unifile format and have them
available for requests levied by the DIA Production
Center. The emphasis of the data collections is on named
objects. Special purpose programs are written to convert
files as they are received.

It is understood that the Unifile system is being
given less attention now than the newly emerging IDHS
(see below).

Comparison with CHIVE: The mechanism of Unifile
and the kind of data applied to it have a striking
resemblance to the CHIVE system. Both systems use the
same type of record structure--linked terms, variable
number of record items. The major difference is that
the CHIVE system includes indexing and data analysis
functions, while Unifile is largely parasitic in its
acquisition of data. This presents severe vocabulary
control problems; the system is at the mercy of the coding
standards established by the contributing organizations.

COMPARISON WITH OTHER SYSTEMS
DIA Systems
1.11.2.1.

-96-

In terms of content, the Unifiles are of the unsynthesized file class defined by CHIVE.

1.11.2.2. Intelligence Data Handling System (IDHS)

The IDHS has been established as a standard data manipulation system to be employed by the intelligence components of the U&S commands. Like Unifile, it is concerned more with a generalized computer mechanism than with the data to be manipulated. The system design has drawn heavily from previous systems in the DOD intelligence environment--SAC, 438L, Fleet Intelligence Centers. It provides generalized capabilities in file definition, file searching and maintenance, and report generation. The language used to specify these functions to the computer is quite rich, but requires considerable training to exploit it properly.

The system is undergoing constant change to increase its power and flexibility. A "user's group" has been established to exchange ideas for improvement and specific data files.

Comparison with CHIVE: As with Unifile, comparison should be limited to the mechanical element of CHIVE.

COMPARISON WITH OTHER SYSTEMS
DIA Systems
1.11.2.2.

Here there are several similarities in functions and capabilities. The logical record structures and the query capabilities are similar. IDHS places more emphasis on generalized methods for defining file input formats and report formats, while CHIVE is concerned with a smaller set of commands which can be mastered as part of the Information Analyst's job. Further, the CHIVE system must deal with information in document index records which is inherently more poorly structured than that of the IDHS files, which are primarily Summary Information Files.

1.11.2.3. DIA Document Retrieval System

The Air Force has employed the Kodak Minicard system for several years as a method for storing and retrieving documents. The systems employed by AF intelligence components were transferred to DIA when it was formed. This system marries the document image and the digital content indexing, referring to it as a single physical unit record. Several pieces of equipment are used to manipulate the unit record "chips" according to programmed logical conditions.

More recently, DIA has acquired several FMA document retrieval systems, which use the same unit record concept

COMPARISON WITH OTHER SYSTEMS
DIA Systems
1.11.2.3.

-98-

as Minicard, except the records are searched in sequence on reels. (See Volume VI.) The FMA systems are being established as a standard for the exchange of document files (and their indexes) among the U&S commands.

Comparison with CHIVE: The purpose and scope of CHIVE as a document retrieval system is significantly different from DIA's. The latter is, in a sense, a switching center for the DOD intelligence network. Relatively shallow indexing is sufficient for DIA because they complement the document retrieval function with a heavy emphasis on information files produced by and for the Unified and Specified commands. Because of the emphasis on in-depth indexing and information extraction in CHIVE, the digital material is best manipulated separately from the document images. In terms of subject indexing, however, each system should be able to benefit from the efforts of the others. The Intelligence Subject Code is an adequate vehicle for the communication of the indexing analysis work that goes into each system. If problems of input time delay can be solved, CHIVE might be assisted by DIA's indexing of its own information reports (which constitute the majority of its document base).

COMPARISON WITH OTHER SYSTEMS
DIA Systems
1.11.2.3.

1.11.3.   AIR FORCE, FTD

The Foreign Technology Division of the Air Force
Systems Command has had two active, large-scale retrieval
efforts--WHITE STORK and CROSS CHECK. The former is the
collateral support document retrieval system for FTD.
It includes open literature and intelligence materials
on aerospace science and technology. It is a manual,
hard copy system where documents are files in several
places (under names, organizations, locations, and
scientific subjects).

The CROSS CHECK system indexes material from Bloc
S&T literature (200 journals). Emphasis is on the topics
listed above. The files on each of these topics are
stored and searched by straightforward computer programs.

More recently, FTD has undertaken a major effort to
consolidate its document processing services with the use
of computer equipment in a system called CIRC (Centralized
Information Retrieval and Control). The objectives of
this system include (a) single-point processing of the
total document data base, (b) a master index and vocabulary,
(c) faster response, (d) user controlled searching and
output screening, and (e) selective notification of newly

                              COMPARISON WITH OTHER SYSTEMS
                              Air Force, FTD
                              1.11.3.

acquired material by means of citations, abstracts, or complete documents based on machine-stored user profiles.

A controlled set of pre-coordinated descriptors is used for indexing. Searches for documents specify descriptors from this set as well as "qualifiers" on document date span, source, etc., which are used to make the search more precise. The degree of filtering of output can also be specified; i.e., any 2, any 3, any 4, etc. descriptors must match. The machine product can be complete index records, citation information, or just document numbers. Recordak Lodestar equipment is used for document image storage and retrieval.

Comparison with CHIVE: The newly developed CIRC system at FTD is quite similar to CHIVE in its mechanical structure and organizational objectives. The principal differences are in the document base and the level of indexing analysis applied to it. In addition, the dissemination features of CIRC have no current parallel in CHIVE.

CHIVE will watch the development of this system with interest. To a large extent, the FTD data base complements rather than duplicates CHIVE. Because of

COMPARISON WITH OTHER SYSTEMS
Air Force FTD
1.11.3.

the demonstrated value of the FTD files to CIA analysts,
the Agency could become a major customer of the new
system. Where appropriate, exchange of data files will
be explored.

1.11.4. NATIONAL PHOTOGRAPHIC INTERPRETATION CENTER

NPIC has computer and Minicard installations to
support its mission planning and photo interpretation
functions. From a central reference standpoint both
facilities are used to provide the photo interpreters with
target information from several sources usually on a
mission-by-mission basis. The Minicard system stores
document images and their indexes. This file consists
primarily of selected PI reports, but also contains
material from other sources of potential use to the
photo interpreter which warrants some in-depth content
indexing. The indexing scheme is based on ██████ STATOTHR

STATOTHR ████████ subject codes. A broader data base is
given header control in a computer-supported index to PI
reports. This index is updated periodically and
disseminated widely in SI and in collateral versions.
The computer is also used to support detailed interpreta-
tion activities through ad hoc assistance of a programming

COMPARISON WITH OTHER SYSTEMS
NPIC
1.11.4.

-102-

staff and direct on-line communication by analysts who can specify a limited set of computations to be performed by the machine.

Comparison with CHIVE: The facilities mentioned above are designed to support a relatively well-defined mission. However, in the area of documentation support, particularly when the problems of selecting appropriate materials for inclusion are considered, NPIC's activities overlap considerably with those proposed by CHIVE. The major differences are in specificity of indexing and physical proximity of the service facilities. OCR is presently supporting NPIC in gathering target material needed in anticipation of a PI requirement. If problems of point of view and adequate response time can be met, the CHIVE system could ultimately handle a significant portion of the NPIC support load.

1.11.5.   NSA

The central reference activity at NSA is comparable in scope and functions with OCR. It services a different user population, of course, but it files similar material in similar categories. Its emphasis is largely on named objects--people, organizations, locations, commodities, but has an extensive coordinate index to technical literature.

COMPARISON WITH OTHER SYSTEMS
NSA
1.11.5.
-103-

Some computer support to the central reference
activity has been undertaken. This has been devoted to
sorting, updating, consolidating, and printing retrieval
aids such as name group tables and gazetteers. The main
files are hard-copy documents which are reproduced and
filed under as many topics as "ticked" off by the
information specialists. Special reference collections
of the standard library type are also maintained.

Comparison with CHIVE: The stated needs of the
NSA user are such that a manual system of direct access
to documents using a single search topic is sufficient.
Searches such as "find references to persons of occupation
A located at B" are difficult to handle in such a system
but must be processed efficiently by CHIVE. Our potential
point of interface with the NSA system will be its use
of indexer and retrieval aids and the vocabulary standards
(such as transliteration standards) which they employ.

COMPARISON WITH OTHER SYSTEMS
NSA
1.11.5.

Appendix 1.A.

BIBLIOGRAPHY OF CHIVE PHASE II PAPERS

1.A.1. WORKING PAPERS

    1. CHIVE/W-3-64, "To Present Alternate Document
Processing Design Systems," 1 March 1964, Unclassified.

    Five alternate document processing schemes are
presented in flow chart form, each accompanied by
a terse description of the process. Purpose is
to evaluate and define the role of individual
support files within the context of each system.

    2. CHIVE/W-4-64, "Support File Task Report #2,"
19 March 1964, Secret.

    Codes, contents and functional use of punched card
geographic location files held by OCR are analyzed
to determine possibility of their use in building
support files.

    3. CHIVE/W-5-64, "Initial CHIVE Input Transcription
Requirements and Techniques," Input Transcription; Initial
Study, 24 March 1964, Unclassified.

    Covers an initial set of design requirements on
input transcription and a list of transcription and
indexing functions. Two potential transcription
techniques are described and the relative advantages
of each discussed.

    4. CHIVE/W-6-64, "First Working Paper on the
Characteristics and Environment of Agency Machine Readable
Data," 24 March 1964, Secret.

    The flow of machine readable (electrically transmitted)
data presently used or manufactured by the Agency is
analyzed. Organizations involved, types of data, and
volumes are discussed.

<div align="right">

BIBLIOGRAPHY
Working Papers
1.A.1.

</div>

5. CHIVE/W-7-64, "Support File Task - Recommendations,"
April 1964, Unclassified.

Specific recommendations are made on the experimental
consolidation of existing location dictionaries. Area
codes, place names and geographic coordinates are
recommended as basic minimum content.

6. CHIVE/W-9-64, "Alternate Document Processing Systems,
8 April 1964, Unclassified.

Additional detail is given on the five alternate
document processing systems presented in CHIVE/W-3-64.
Files are defined and the roles of the indexer,
analyst, and dictionary editor discussed.

7. CHIVE/W-10-64, "CHIVE Indexing Concept - Analyst
Survey," April 1964.

An explanation of the CHIVE indexing concept with
illustrations of CHIVE indexing notions. A paper
prepared for transmittal to production shop analysts
to obtain their reactions to CHIVE proposals.

8. CHIVE/W-11-64, "Second and Final Working Paper on the
Exploitation of Agency Machine Readable Data," 7 May 1964,
Secret.

This paper covers message handling, information
extraction, indexing, dissemination, and storage
of the three major MR categories listed in W-6-64.

9. CHIVE/W-12-64, "Current Document Dissemination
Procedures," 12 May 1964, Secret.

Provides a detailed description of major dissemination
groups, manpower, document types and volumes, dissemina-
tion practices and criteria, and general document flow.

10. CHIVE/W-13-64, "Preliminary Economic Assessment of
Alternate Means for Input Transcription," 15 May 1964.

Cost comparisons for input transcription are made of
punched cards, punched paper tape, optical page
readers and CRT consoles. Results are indecisive and
further study recommended. See also W-5-64.

<div align="right">
BIBLIOGRAPHY
Working Papers
1.A.1.
</div>

SECRET

11. CHIVE/W-14-64, "Security Classification Control and Compartmentation," 8 June 1964, Secret.

Security classification, dissemination control, and compartmentation are defined and discussed briefly. Recommendations are made as to the processing of such data in the CHIVE system. Pertinent regulations are carried as enclosures.

12. CHIVE/W-15-64, "Reference Aids Used to Support OCR Input Processing," 12 June 1964, Secret.

Reference aids used by OCR Divisions are charted by Division. Standard works are listed separately from reference aids created from machine record holdings of the divisions.

13. CHIVE/W-16-64, "General Functional Characteristics of the Executive Control Subsystem," 19 June 1964, Confidential.

Executive Control is defined in terms of its relation to operational program control. Ten executive control functions are listed and the program tasks necessary to execution of these functions are listed.

14. CHIVE/W-17-64, "Performance Specifications for: 1) Reference Subsystem; 2) Information Subsystem 3) Document Delivery Subsystem," 22 June 1964, Confidential.

Summarizes and updates performance specs for the three subsystems originally outlined in CHIVE/R-1-64.

15. CHIVE/W-18-64, "Results of Analyst Survey Task (Phase I)," 25 June 1964, Secret/No Foreign Dissemination.

Paper includes a survey package, a list of survey respondents arranged by organization, and a compilation and discussion of survey findings.

16. CHIVE/W-19-64, "Comparison of Generalized Intelligence Data Handling Computer Programs," 8 July 1964, Secret.

Compares 1410 F.F.S. ██████████████ **25X1A**
Phase I, 7090 F.F.S. (IDHS) and UNIFILE. Describes

BIBLIOGRAPHY
Working Papers
1.A.1.

-107-

SECRET

each system in terms of data organization, file building and maintenance, and retrieval and output. Record formats are shown for each system.

17. CHIVE/W-21-64, "Functional Characteristics of the Reference Subsystem," 17 July 1964, Confidential.

A discussion of the header and content indexing techniques proposed for the CHIVE Document Storage and Retrieval System. Retrieval parameters are listed for Header Data, Personalities, Organization/ Facilities, Meetings/Conferences and Locations. Residual file design and authority files are also discussed.

18. CHIVE/W-22-64, "Information Subsystem: Functional Characteristics," 17 July 1964, Secret.

A description of major file types, alternative methods for creating these records, error checking, file design, analyst file communication, file processing, query processing, file conversion and hardware considerations.

19. CHIVE/W-23-64, "Functional Characteristics of the Document Delivery Subsystem," 28 July 1964, Confidential.

This subsystem is discussed from three points of view including input processing and control, storage characteristics and problems, and output or document delivery. Conversion of existing files is discussed briefly.

20. CHIVE/W-24-64, "Initial Indexing Experiment," 27 July 1964, Confidential/No Foreign Dissemination.

A brief analysis of the results obtained from indexing 500 collateral documents. The indexing system is explained and a tagged list of CHIVE retrieval parameters appended.

BIBLIOGRAPHY
Working Papers
1.A.1.

CONFIDENTIAL

21.   CHIVE/W-26-64, "Personnel and Management Subsystem – Background and General Specifications," 14 August 1964, Confidential.

A reiteration of CHIVE concepts and objectives followed by a discussion of alternative organizational configurations through which these objectives might be attained. An area approach is selected and ███████ recommended as the initial point of entry.

25X6

22.   CHIVE/W-27-64, "Header Data Transcription Task," 20 August 1964, Secret.

Describes the capturing of bibliographic descriptions of documents by clerical personnel. A procedure and instruction manual is attached.

23.   CHIVE/W-28-64, "Preliminary Functional Design of the Executive Control Subsystem," 21 August 1964, Unclassified.

A continuation and expansion of CHIVE/W-16-64. Discussions are on basic control functions and are intended as a framework for more intensive design work.

24.   CHIVE/W-29-64, "Preliminary Functional Design of the Information Subsystem," 1 September 1964, Secret.

Describes the Information Subsystem as a generalized formatted file system capable of storing and maintaining many different formats of data files and of retrieving all or selected portions of these files. The paper summarizes the functions of the subsystem, appraises file content, and describes methods of establishing maintaining and using the formatted files.

25.   CHIVE/W-30-64, "OCR/CHIVE Indexing Experiment Guide," 30 September 1964, Confidential.

An introduction to the CHIVE Indexing Concept and to the Fall experiment. The Guide is basically an indexing manual.

BIBLIOGRAPHY
Working Papers
1.A.1.

26. CHIVE/W-31-64, "Summary of Statistical Data Extracted from CHIVE Documentation," 27 October 1964, Secret.

A collection of data taken from CHIVE documentation published to date. Figures include projected CHIVE system performance specs, document input rates, ████████ inputs and requests, etc.

25X6

27. CHIVE/W-32-64, "Survey of Physical Characteristics of CHIVE Documents," 10 November 1964, Secret.

Provides a number of figures on volume of document receipts by source including annual volumes, projected annual page volumes, document dimensions, front and back printing, etc.

28. CHIVE/W-33-64, "CHIVE Indexing Test - Quantitative Analysis," 10 November 1964, Secret.

This paper explains the methodology for measuring indexer reliability and query recall and relevancy.

29. CHIVE/W-1-65, "Functional Design of the Reference Subsystem," 8 January 1965, Unclassified.

Specifies alternate designs for a CHIVE EDP document reference subsystem. Includes notes leading to a design for a reference language and program implementation.

1.A.2.   MEMORANDA

30. CHIVE/C-5-64, "CHIVE Indexing Approach and Attribute Selection," 7 February 1964, Confidential.

Details the rationale behind the CHIVE named object indexing concept and suggests criteria for selecting attributes of named objects for index control.

BIBLIOGRAPHY
Memoranda
1.A.2.

31. CHIVE/C-15-64, "CHIVE Evaluation Conference - Basic Guidelines for CHIVE Outlined," 29 April 1964, Confidential.

Summary of CHIVE goals, design approach and initial system characteristics. A proposed time schedule for CHIVE is presented.

32. CHIVE/C-17-64, "Survey of Agency Pneumatic Tube System," 11 May 1964, Confidential.

A description of the system.

33. CHIVE/C-18-64, "CHIVE File Definitions," 13 May 1964, Confidential.

Broadly defines document references and information files. The latter are further defined to include CHIVE support files (indexer aids), converted OCR machine records and general purpose and special project information files.

34. CHIVE/C-19-64, "Current Map Library Activities," 8 May 1964, Secret.

A brief discussion of Map Library organization, holdings, activities, and some potential EDP applications.

35. CHIVE/C-20-64, "Functional Activities and Holdings of the Map Library," 28 May 1964, Secret.

A discussion of current map processing activity. This report covers procurement, indexing, dissemination, storage, and retrieval functions of the Map Library.

36. CHIVE/C-21-64, "OCR Document Request Statistics-- 'Over-the-Counter' Requests vs. Subject Searches," 8 June 1964, Confidential.

Comparative data on documents requested by number and those identified by subject search. Figures given for both Intellofax and SR systems.

BIBLIOGRAPHY
Memoranda
1.A.2.

37. CHIVE/C-24-64, "OCR/Chive Indexing Experiment, Fall 1964," 7 August 1964, Confidential.

Paper describes the objectives, planning, and methodology of the Fall experiment. Personnel assignments, indexing tools, the document base, selection, tests and test analysis are covered in some detail.

38. CHIVE/C-28-64, "Intelligence Files: Requirements, Characteristics, Problems," 30 September 1964, Confidential.

A bibliography of reports and papers and a list of CIA personnel familiar with various aspects of the subject.

39. CHIVE/C-29-64, "Alternative Approaches to the Handling of Maps," 8 October 1964, Confidential.

Integration of the Map Library with CHIVE Geographic divisions, centralized indexes/decentralized processing, retention of present configuration, Map Library as a distinct entity within CHIVE and retention of the present configuration with semi-duplicative processing are discussed.

40. CHIVE/C-Unnumbered, "Implications of CHIVE/R-1-63 from the OCR Viewpoint," 10 April 1964, Confidential.

Paper reflects the impressions and opinions of the CHIVE Support Staff based on their reading of CHIVE/R-1-63. All major CHIVE concepts from 1-63 are discussed and their impact on OCR assessed.

1.A.3. REPORTS

41. CHIVE/R-1-63, "Preliminary System Design Report," 1 December 1963, Secret/No Foreign Dissemination/Continued Control.

This paper presents a brief background on the CHIVE

BIBLIOGRAPHY
Reports
1.A.3.

concept, discusses system objectives and lays down preliminary functional specifications. The report was written as a vehicle to guide CHIVE design activity.

42. CHIVE/R-1-64, "CHIVE Evaluation Conference (20-30 April 1964): Notes, Working Papers, Decisions," 4 May 1964, Secret.

25X1A

An unedited assembly of working papers prepared during an eight-day conference at ▮▮▮▮ Report covers System Objectives, Personnel and Management, CHIVE System Performance Specifications, Input System Design, Task Requirements, Tasks, and Assignments.

43. CHIVE/R-2-64, "A Comparative Analysis of Document Delivery Systems for Large and Active Files," August 1964, Administrative Internal Use Only.

This report analyzes eleven candidate storage and delivery systems in terms of criteria such as cost, staff, space and established requirements.

44. CHIVE/R-1-65, "A Comparative Analysis of Document Delivery Systems," 1 March 1965, Unclassified.

Presents an evaluation of currently available systems for document image storage and retrieval. Systems evaluated on cost, space, and manning.

45. CHIVE/R-2-65, "A Comparative Analysis of Input, Transcription Techniques," 1 March 1965, Unclassified.

Presents an evaluation of four basic methods for transcribing formatted data into machine form-- card punching, paper tape punching, on-line terminal, and page reader.

BIBLIOGRAPHY
Reports
1.A.3.

1.A.4.   MISCELLANEOUS PAPERS

46.   IBM, "Document Storage and Retrieval System" Proposal from IBM Data Processing Division, Washington, D. C., 19 February 1965, Unclassified, IBM Proprietary.

47.   Magnavox Company, "MAGNAVUE Document Storage and Retrieval System," MRL Proposal No. TP-1315, 15 February 1965, Unclassified, Magnavox Proprietary.

48.   OCS, "Document Numbering Systems," Memorandum for the Record, 21 August 1964, Secret.

A collection of charts and memoranda on numbering systems in use today. Memo refers reader to Special (SI) Numbering System. This paper should be read in conjunction with W-27 and with the Header Data Transcription Manual.

BIBLIOGRAPHY
Miscellaneous Papers
1.A.4.